

**UNIVERSIDAD AUTÓNOMA DE MADRID**

**FACULTAD DE PSICOLOGÍA**

**Departamento de Psicología Social y Metodología**



**TESIS DOCTORAL**

**EVALUACIÓN DE LA ESTIMACIÓN DE LA FIABILIDAD EN CONDICIONES DE DATOS ASIMÉTRICOS,  
CONGENÉRICOS, CATEGÓRICOS Y EN PRESENCIA DE MULTIDIMENSIONALIDAD**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR PRESENTADA POR:**

**Italo Pietro Lucco Trizano Hermosilla**

Director

**Jesús María Alvarado Izquierdo**

Madrid, 2017



Dedicada a la persona más importante de mi vida: mi mamá.

## **Agradecimientos**

En primer lugar quiero agradecer a mi director Dr. Jesús María Alvarado Izquierdo por su constante apoyo e infinita paciencia en este proceso de tesis, sus valiosos aportes teóricos y mirada crítica permitieron dar forma y enfocar el trabajo.

Debo dar gracias a CONICYT que, por medio de su programa Becas Chile Doctorado en el Extranjero, permitió financiar esta etapa de estudios doctorales.

Además debo agradecer al Dr. Vicente Ponsoda por todas las gestiones realizadas en su rol de tutor.

Finalmente, agradezco a mi madre, Pamela Hermosilla, ya que sin ella nada de esto sería posible.

# ÍNDICES

## Índice de Contenido

ÍNDICES.....	5
ÍNDICES DE TABLAS .....	7
ÍNDICE DE FIGURAS .....	8
RESUMEN .....	9
INTRODUCCIÓN .....	11
ESTIMACIÓN DE LA FIABILIDAD DESDE LA TEORÍA CLÁSICA DE LOS TEST .....	15
Coeficiente alfa de Cronbach .....	18
Supuestos del coeficiente alfa.....	21
ESTIMACIÓN DE LA FIABILIDAD DESDE LOS MODELOS DE ECUACIONES ESTRUCTURALES .....	31
Familia de coeficientes omega .....	33
Estimación de la fiabilidad desde el modelo bifactor .....	36
GREATEST LOWER BOUND .....	40
ESTIMACIÓN DE LA FIABILIDAD EN PRESENCIA DE DATOS CATEGÓRICOS.....	42
Coeficiente alfa ordinal .....	46
Coeficiente omega ordinal .....	47
OBJETIVOS DE INVESTIGACIÓN .....	48
ESTUDIO 1: ESTIMACIONES DE LA FIABILIDAD EN MODELOS UNIDIMENSIONALES .....	49
ESTUDIO 1.1: ESTIMACIONES DE LA FIABILIDAD CON DATOS CONTINUOS .....	49
Método.....	49
Resultados .....	51
Conclusión .....	54
ESTUDIO 1.2: ESTIMACIÓN DE LA FIABILIDAD CON DATOS CATEGÓRICOS.....	55
Método.....	55
Resultados .....	58
Conclusión .....	69
ESTUDIO 2: ESTIMACIÓN DE LA FIABILIDAD EN MODELOS MULTIDIMENSIONALES.....	71
EJEMPLOS DE ESTIMACIÓN DE LA FIABILIDAD CON CORRELACIÓN ENTRE ERRORES. ....	72
Caso 1. Modelo Unidimensional de 5 ítems.....	72
Caso 2: 6 ítems con correlación entre errores .....	73

Caso 3. 9 ítems y 3 factores específicos: .....	74
ESTUDIO 2.1: SESGOS EN LA ESTIMACIÓN DE LA FIABILIDAD EN MODELOS BIFACTOR.....	77
Método.....	77
Resultados .....	80
Conclusión .....	83
ESTUDIO 2.2: SESGOS EN LA ESTIMACIÓN DE LA FIABILIDAD EN PRESENCIA DE ÍTEMS CON INFORMACIÓN REDUNDANTE.....	84
Método.....	84
Resultados .....	85
Conclusión .....	90
ESTUDIO 3: ESTIMACIONES DE LA FIABILIDAD CON DATOS EMPÍRICOS .....	93
ESTUDIO 3.1: RE-ANÁLISIS DE LAS EVIDENCIAS DE FIABILIDAD DEL CECSC. ....	93
Método.....	93
Resultados .....	96
Conclusión .....	100
ESTUDIO 3.2: RE-ANÁLISIS DE LAS EVIDENCIAS DE FIABILIDAD DEL AAI-A.....	101
Método.....	101
Resultados .....	103
Conclusión .....	108
DISCUSIÓN Y CONCLUSIONES FINALES. ....	111
1. ESTIMACIÓN DE LA FIABILIDAD EN MODELOS UNIDIMENSIONALES.....	111
2. ESTIMACIÓN DE LA FIABILIDAD EN MODELOS MULTIDIMENSIONALES. ....	115
3 CONCLUSIONES GENERALES .....	118
REFERENCIAS.....	123

## ÍNDICES DE TABLAS

Tabla 1: RMSE y Sesgo para las condiciones de tau-equivalencia y datos congénéricos para 6 ítems, tres tamaños muestrales y diferente número de ítems asimétricos.....	52
Tabla 2: RMSE y Sesgo para las condiciones de tau-equivalencia y datos congénéricos para 6 ítems, tres tamaños muestrales y diferente número de ítems asimétricos.....	53
Tabla 3: 2 categorías 6 ítems congénérico (fiabilidad = .731) .....	59
Tabla 4: RMSE y Sesgo de los coeficientes de fiabilidad para la condición de 2 categorías .....	60
Tabla 5: 3 categorías 6 ítems congénérico (fiabilidad = .731) .....	62
Tabla 6: RMSE y Sesgo de los coeficientes de fiabilidad para la condición de 3 categorías .....	62
Tabla 7: 4 categorías 6 ítems congénéricos (fiabilidad = .731).....	64
Tabla 8: RMSE y Sesgo de los coeficientes de fiabilidad para la condición de 4 categorías .....	65
Tabla 9: 5 categorías 6 ítems congénéricos (fiabilidad = .731).....	67
Tabla 10: RMSE y Sesgo de los coeficientes de fiabilidad para la condición de 5 categorías .....	68
Tabla 11: Matriz de correlación entre 5 ítems.....	73
Tabla 12: Matriz de correlación entre 6 ítems con correlación entre errores de ítem 5 y 6 .....	74
Tabla 13: Matriz de correlación entre 9 ítems con múltiples correlaciones entre errores.....	75
Tabla 14: Recuperación de la fiabilidad con diferentes estimadores .....	81
Tabla 15: Sesgos en la estimación de la fiabilidad en modelos bifactor .....	82
Tabla 16: Estimaciones de la fiabilidad del factor general obtenidas para las condiciones simuladas .....	87
Tabla 17: Sesgo de los coeficientes de fiabilidad en condiciones de ítem redundantes.....	89
Tabla 18: Estadísticos descriptivos de los ítems del CECSCCE .....	96
Tabla 19: Ajuste de los modelos CECSCCE .....	97
Tabla 20: Diferencia entre los modelos mediante el factor de corrección para MLM o mediante DIFFTEST para WLSMV .....	98
Tabla 21: Saturaciones factoriales de los modelos contrastados.....	99
Tabla 22: Evidencias de fiabilidad para CECSCCE basado en M2.....	100
Tabla 23: Estadísticos descriptivos de los ítems del AAI-A .....	104
Tabla 24: Ajuste de los modelos para la escala AAIA .....	105
Tabla 25: Diferencia entre los modelos mediante el factor de corrección para MLM o mediante DIFFTEST para WLSMV .....	106
Tabla 26 : Saturaciones con el Modelo Bifactor .....	107
Tabla 27: Evidencias de fiabilidad para AAI-A basado en M3.....	108

## ÍNDICE DE FIGURAS

Figura 1: Representación de los modelos (a) Tau-equivalente y (b) Congenéricos con 6 ítems. ....	25
Figura 2: Ejemplo de un modelo unidimensional con correlaciones entre los errores de los ítems 1 y 2 y los ítems 5 y 6. ....	28
Figura 3: Representación de un Modelo Bifactor considerando 18 ítems y una estructura con un Factor General, común a todos los ítems, y 3 factores grupales cada uno compuesto por 6 ítems...	37
Figura 4: Ejemplos de ítems asimétricos con dos, tres, cuatro y cinco categorías de respuesta. ....	56
Figura 5: Modelo Bifactor para 9 ítems con 6 ítems redundantes.....	76
Figura 6: Ejemplo de modelo Bifactor poblacional simulado para la condición de fiabilidad = .923 .....	79
Figura 7: Estimaciones de la fiabilidad para el modelo Bifactor .....	83
Figura 8: Gráfico del modelo Bifactor con datos categóricos para AAI-A. ....	107



## RESUMEN

La adecuada estimación de la fiabilidad de las medidas que se utilizan en psicología es un elemento fundamental para juzgar tanto su calidad como su potencial como instrumento diagnóstico y pronóstico. A partir de la teoría clásica de los test y su extensión factorial hacia los modelos de variables latentes, junto a la aplicación de los modelos de ecuaciones estructurales, se han desarrollado diferentes procedimientos de estimación que han dado lugar a un conjunto de coeficientes que literalmente cubren todo el alfabeto griego desde alfa hasta omega. De todos los coeficientes destaca por su amplia difusión y utilización el coeficiente de consistencia interna alfa de Cronbach, un excelente estimador de la fiabilidad cuando se cumplen los supuestos del modelo. Sin embargo, en la práctica difícilmente se cumplen algunos o varios de los supuestos (e.g., tau-equivalencia, normalidad, escalamiento intervalar, unidimensionalidad o ausencia de correlación entre errores) y cuando esto ocurre tanto el coeficiente alfa como la mayoría de los coeficientes propuestos muestran problemas más o menos severos en algunos casos de subestimación (e.g., ante distribuciones asimétricas y categóricas) o más preocupante de sobreestimación (e.g., correlaciones entre errores o estructuras multidimensionales). En la presente memoria se revisan cada uno de estos problemas y se analizan mediante técnicas de simulación tres escenarios habituales en investigación aplicada: a) estimación en datos unidimensionales con incumplimiento del supuesto de normalidad y/o tau-equivalencia, b) estimación en datos unidimensionales ordinales y congéntricos con diferente número de categorías y c) estimación en datos multidimensionales o con presencia de ítems redundantes (errores positivamente correlacionados). Para evaluar estos escenarios se realizaron cuatro estudios de simulación Montecarlo y dos estudios con datos reales empíricos en los que se comparan los procedimientos alfa de Cronbach, omega total, omega jerárquico, tanto en su versión lineal

como ordinal, GLB y GLB algebraico con el objetivo de identificar aquellos que muestran ser más robustos a los distintos incumplimientos de los supuestos del modelo clásico. La discusión se centra en los estimadores que presentan mejor desempeño para las condiciones evaluadas y la importancia de evaluar la distribución de los datos, sus características y la estructura interna (mediante técnicas factoriales y aplicación del modelo bifactor), como requisito para obtener evidencias adecuadas de la fiabilidad de los datos, ya que de no hacerlo se pueden obtener resultados sesgados, como concluir equivocadamente que la puntuación global del test presenta una alta fiabilidad, cuando la realidad es que solo factores grupales que lo conforman presentan altas fiabilidades, en consecuencia la medición del factor común (a veces inexistente) conlleva un alto grado de error, a pesar de lo que pudieran indicar los coeficientes de fiabilidad habituales, ya que generalmente no son sensibles al problema de la multidimensionalidad.

## INTRODUCCIÓN

Los procesos de medición en cualquier disciplina científica se encuentran afectados por un componente de error (Carmines & Zeller, 1979; Nunnally & Bernstein, 1995), la psicología no es ajena a esta situación, muy por el contrario, la presencia de errores se encuentra íntimamente relacionada con la naturaleza latente de las variables psicológicas (Fornell & Larcker, 1981; Kane, 2008; Muñiz, 1998) tales como la inteligencia, la ansiedad, la depresión, la habilidad de tocar un instrumento musical, la cultura, etc., y con las herramientas tecnológicas (e.g., test, cuestionarios, escalas, etc.) que se han desarrollado para evaluar estos constructos (Muñiz, 2000; Muñiz & Fonseca-Pedrero, 2008; Schmidt & Hunter, 1996). Pese a estas dificultades, la teoría psicométrica ha intentado dar respuesta al problema del error de medida.

Tal y como su nombre sugiere, la naturaleza latente de los constructos psicológicos hace necesario elaborar modelos teóricos en los que se especifique el efecto que una variable latente puede tener sobre variables observadas o manifiestas (Jöreskog & Sorbom, 1979). En este contexto se entiende que, por ejemplo, el ítem de un test corresponde a la variable observada, también llamado indicador, de un constructo subyacente comúnmente denominado factor (Bollen, 2002). Así, la respuesta del individuo en ese ítem dependerá, en parte, del nivel de rasgo en el o los factores y, en parte, del error de medición. Conviene comprender entonces que los errores no son otra cosa que todo aquello que escapa a la explicación de nuestro modelo (Kane, 2008).

El error de medida aleatorio corresponde a fluctuaciones impredecibles, factores transitorios e irrepetibles, ruido aleatorio, etc., que afecta a la precisión de las medidas en el momento en que éstas se realizan; a nivel teórico su esperanza matemática es 0. Por ejemplo, un estudiante al ser evaluado en un test de conocimiento puede estar más distraído en una aplicación y más concentrado en la siguiente, o bien estar más nervioso o más calmado, también pueden existir situaciones o factores externos (el horario de aplicación del test, un ruido de sirena a la distancia, etc.) todo lo cual generará fluctuaciones en la puntuación de ese test que variarán de una aplicación a otra.

De este modo, los procesos de medición realizados en ciencias sociales, educación y psicología deben mostrar evidencias de dos propiedades fundamentales: la validez y la fiabilidad. La primera ha sufrido diversas modificaciones en su definición a lo largo de la historia (Elosua, 2003; Prieto & Delgado, 2010), sin embargo mantiene su rol como la propiedad más importante de todas. Actualmente la validez es entendida como el grado en que la evidencia empírica y el sustento teórico permiten interpretar las puntuaciones de un test bajo un modelo determinado (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Complementariamente, la fiabilidad se ocupa de la precisión de las puntuaciones obtenidas con el test, es decir, en qué grado las puntuaciones se encuentran libres de error de medida. Por lo tanto, la relación sustantiva entre validez y fiabilidad se encuentra en que mientras la primera busca responder a la pregunta “¿Qué estamos midiendo?”, la segunda lo complementa evaluando “¿Con qué grado de error lo estamos midiendo?”.

En otras palabras, los procesos de medición en ciencias sociales y en psicología deben garantizar que las puntuaciones obtenidas sean un indicador fiable, preciso, de un constructo adecuadamente representado (Deng & Chan, 2016). Es importante considerar que estas dos propiedades, validez y fiabilidad, no son entendidas como características inmutables y sempiternas de los test, su conceptualización actual indica que son propiedades que se aplican a las puntuaciones de los test y que, por lo tanto, pueden variar de aplicación en aplicación (Thompson & Vacha-Haase, 2000; Wilkinson & American Psychological Association Task Force on Statistical Inference, 1999).

A modo general, para el análisis y evaluación del error de medida, se han sugerido tres procedimientos que permiten estimar el coeficiente de fiabilidad de las puntuaciones de un test (Santisteban & Alvarado, 2001): a) análisis de consistencia interna; b) test-retest o estabilidad temporal; y c) formas paralelas. De todos ellos, el primero es el más utilizado en la práctica por los investigadores (Hogan, Benjami, & Brezinski, 2000) y es del que nos ocuparemos en esta memoria, debido a que, entre otras razones, tiene la ventaja de requerir una sola aplicación del instrumento (Abad, Olea, Ponsoda, & García, 2011). Actualmente, autores como Green y Yang (2009a) y Crutzen y Peters (2015) sostienen que previo a evaluar la fiabilidad de las puntuaciones de los instrumentos, por cualquiera de estos procedimientos, es necesario comprobar tanto la estructura empírica subyacente a los datos, como el cumplimiento de los supuestos de los estimadores de fiabilidad, para ello proponen modelizar mediante ecuaciones estructurales las relaciones entre el conjunto de ítems y el o los factores latentes, permitiendo contrastar los planteamientos teóricos con los datos empíricos. Así, por ejemplo, antes de evaluar la precisión de las puntuaciones de un test, se deben ofrecer evidencias de que la estructura empírica obtenida es compatible tanto con la teoría subyacente

como con los estudios previos, solo entonces se procederá a realizar las estimaciones con los coeficientes más adecuados atendiendo a la naturaleza de la estructura planteada (e.g: modelos unidimensionales, tau-equivalentes, con correlación entre errores, jerárquicos de segundo orden, Bifactor, de medidas repetidas, de invarianza de parámetros, etc.).

En una interpretación general e intuitiva del coeficiente de fiabilidad, que corresponde a un indicador numérico con valores teóricos comprendidos entre 0 y 1; los valores cercanos a 0 indican que la puntuación del instrumento solo está midiendo ruido, en tanto los valores cercanos a 1 dan cuenta de que las puntuaciones son casi perfectas, indicando que la varianza del error de medida es pequeña (Muñiz, 1998). En este continuo, la pregunta “¿Cuál es el valor óptimo de fiabilidad que requieren las puntuaciones del test?” no tiene una respuesta fácil y unánime por parte de la literatura. Por ejemplo, Nunnally (1978) sugiere como valor aceptable, en las etapas tempranas de investigación, una fiabilidad en torno a .70, mientras que Carmines & Zeller (1979) proponen un valor más exigente de .80. Sin embargo, contrariamente a la intuición, un valor excesivamente alto del coeficiente de fiabilidad (cercano a 1) se puede deber a redundancia en los ítems (preguntar varias veces lo mismo de formas distintas) por lo que Streiner (2003) considera que el valor del coeficiente de fiabilidad no debe ser mayor a .90. Así, en la actualidad el planteamiento es que el nivel de fiabilidad deseable se encuentra estrechamente ligado a las consecuencias que se puedan derivar de la puntuación obtenida por el sujeto en un determinado test. Por lo tanto, si las consecuencias son de bajo impacto para la vida del sujeto (como en áreas de investigación básica) se puede considerar aceptables niveles moderados de fiabilidad. Sin embargo, a medida que las consecuencias que se puedan derivar de la puntuación obtenida por un sujeto tengan mayor impacto en su vida (e.g., pruebas psicomotrices de conducción, test de

selección de personal, pruebas de ingreso a la universidad, etc.), los niveles de fiabilidad deberán ser más exigentes (Lance, Butts, & Michels, 2006; Nunnally & Bernstein, 1995).

A continuación se presentará en primer lugar la Teoría Clásica de los Test (TCT), como uno de los *framework* psicométricos que se ha planteado la cuestión del error en la medición, así también se describirá brevemente el coeficiente Alfa ( $\alpha$ ) de Cronbach y sus supuestos, destacando las consecuencias que el incumplimiento de ellos tiene en la estimación de la fiabilidad. Posteriormente se presentará el *framework* del Análisis Factorial, el que se puede entender como una extensión de la TCT, dentro de los Modelos de Ecuaciones Estructurales, y sus ventajas para la evaluación de la precisión al trabajar con variables latentes, describiendo la familia de coeficientes Omega ( $\omega$ ), adecuados a distintas situaciones e introduciendo los principios del modelo Bifactor. A continuación se explorará otro coeficiente denominados *Greatest Lower Bound* (GLB) como estimador de la fiabilidad. Finalmente se expondrá el efecto que tienen los indicadores categóricos, dicotómicos o politómicos, y los niveles de asimetría en la estimación de la fiabilidad, presentando las alternativas que se han propuesto para el trabajo con este tipo de variables.

## ESTIMACIÓN DE LA FIABILIDAD DESDE LA TEORÍA CLÁSICA DE LOS TEST

Propuesta hace más de un siglo por Spearman (1904), posteriormente formalizada por Gulliksen (1950) y rigurosamente sistematizada más tarde tanto por Lord y Novick (1968) como por Zimmerman (1975) la TCT es uno de los primeros esfuerzos por desarrollar un modelo matemático que sustentase la medición en psicología modelizando las respuestas de

los sujetos. A continuación se presentarán de manera introductoria los principales aspectos de la TCT. Para profundizar en la misma puede consultarse cualquiera de los manuales disponibles tanto en español (Abad et al., 2011; Martínez-Arias, Hernández, & Hernández, 2006; Martínez-Arias, 1995; Muñiz, 2000; Santisteban, 2009) como en inglés (Lord & Novick, 1968; Nunnally & Bernstein, 1995; Raykov & Marcoulides, 2011). Tal como se ha mencionado previamente, la TCT parte de la idea de que las puntuaciones observadas se encuentran afectadas por dos componentes, la puntuación verdadera y el error de medida, lo anterior es formalizado con la sencilla ecuación:

$$X_{ij} = V_{ij} + e_{ij} \quad (1)$$

Se entiende que la puntuación observada de un sujeto  $i$  en un Test  $j$  ( $X_{ij}$ ) se compone de dos partes, la puntuación verdadera de esa persona en el Test ( $V_{ij}$ ) y un error ( $e_{ij}$ ); de esta ecuación se deducen los siguientes supuestos inherentes al modelo: a) los errores de medida son independientes unos de otros [ $Cov(e_j, e_k) = 0$ ], b) no existe correlación entre el error y la puntuación verdadera [ $Cov(e, V) = 0$ ], c) la esperanza matemática de los errores es igual a cero [ $E(e) = 0$ ] y d) la esperanza matemática de la puntuación observada es igual a la puntuación verdadera [ $E(X) = V$ ]. No se debe perder de vista que estos supuestos se plantean en el caso de que un test se administrara infinitamente a un mismo sujeto, de aquí que la esperanza del error se anule y la puntuación verdadera sea la esperanza de la puntuación observada. Por lo general, en la realidad tenemos una sola aplicación del instrumento en una muestra concreta, si asumimos que es representativa de la población y que en esta hay un gran número de sujetos con iguales puntuaciones verdaderas se puede mantener que la puntuación verdadera es la esperanza matemática de la puntuación observadas en todos los



sujetos con igual puntuación verdadera. De este supuesto (ver Lord y Novick, 1968) se derivan las siguientes relaciones.

Siguiendo los planteamientos de la TCT, la varianza de la puntuación observada ( $\sigma^2_X$ ), se compone, en parte por varianza atribuida a la puntuación verdadera ( $\sigma^2_V$ ) y, en parte, por varianza del error ( $\sigma^2_e$ ), su expresión formal es:

$$\sigma_X^2 = \sigma_V^2 + \sigma_e^2 \quad (2)$$

Además la relación entre la covarianza de las puntuaciones empíricas y la covarianza de las puntuaciones verdaderas es igual a la varianza de las puntuaciones verdaderas:

$$Cov(X, V) = \sigma_V^2 \quad (3)$$

En tanto, la covarianza entre las puntuaciones observadas y los errores de medida corresponde a la varianza de dichos errores:

$$Cov(X, e) = \sigma_e^2 \quad (4)$$

Así, el índice de fiabilidad ( $\rho_{XV}$ ) puede ser expresado formalmente como:

$$\rho_{XV} = Cor_{(X,V)} = \frac{Cov(X,V)}{\sqrt{\sigma_X^2 \sigma_V^2}} = \frac{Cov((V+e),V)}{\sqrt{\sigma_X^2 \sigma_V^2}} = \frac{\sigma_V^2}{\sqrt{\sigma_X^2 \sigma_V^2}} = \sqrt{\frac{\sigma_V^2}{\sigma_X^2}} = \frac{\sigma_V}{\sigma_X} = 1 - \frac{\sigma_e}{\sigma_X} \quad (5)$$

Por su parte, el coeficiente de fiabilidad ( $\rho_X$ ) es equivalente al cuadrado del índice de fiabilidad ( $\rho_X = \rho_{XV}^2$ ), por lo tanto, mientras el índice de fiabilidad se puede interpretar como la correlación entre la puntuación total y la puntuación verdadera, el coeficiente de fiabilidad

se puede entender, por un lado, como un coeficiente de determinación ( $R^2$ ) que permite explicar el grado de relación lineal entre la puntuación observada y la puntuación verdadera y, por el otro, como la razón entre la varianza verdadera y la varianza total de las puntuaciones observadas:

$$\rho_X = \frac{\sigma_V^2}{\sigma_X^2} = 1 - \frac{\sigma_e^2}{\sigma_X^2} \quad (6)$$

Sin embargo, con las ecuaciones anteriormente presentadas no es posible encontrar empíricamente el valor del coeficiente de fiabilidad debido a que la puntuación verdadera no es directamente observable. Un procedimiento, desde este marco, que se ha desarrollado para estimar la fiabilidad de las puntuaciones de un test en una única aplicación es el coeficiente  $\alpha$  de Cronbach.

### **Coeficiente alfa de Cronbach**

El coeficiente  $\alpha$  es probablemente el procedimiento más utilizados para evaluar la fiabilidad de las puntuaciones de los instrumentos de medida en ciencias sociales (Cho & Kim, 2015; Green & Yang, 2009b; Hogan et al., 2000; Sijtsma, 2009a; Zumbo & Rupp, 2004). A pesar de que su autoría se atribuye normalmente a Cronbach (1951), una versión equivalente al coeficiente  $\alpha$  fue propuesto con anterioridad por Guttman (1945), que lo denominó como lambda 3 ( $\lambda_3$ ), en este sentido, también resulta equivalente al presentado por Hoyt (1941) desde la perspectiva del análisis de varianza, e incluso se puede encontrar su origen en la famosa fórmula KR-20 (Kuder & Richardson, 1937) para datos dicotómicos. El coeficiente  $\alpha$  se ha considerado como equivalente al promedio de todas las estimaciones de fiabilidad mediante el procedimiento de *Split-Half* (estimación mediante dos mitades de la escala). El

procedimiento  $\alpha$  se encuentra como opción por defecto al estimar la fiabilidad por consistencia interna en muchos de los *softwares* más utilizados para el análisis de datos, lo cual ha contribuido a difundir su uso. Formalmente este coeficiente se puede expresar en el numerador como la sumatoria de varianzas de los  $i$  ítems o como sumatoria de covarianza entre todos los pares de ítems  $(i,j)$ :

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum \sigma_{Xi}^2}{\sigma_X^2} \right) = \frac{k}{k-1} \left( \frac{\sum_{i \neq j} Cov(X_i, X_j)}{\sigma_X^2} \right) \quad (7)$$

Donde  $k$  es el número de ítems de la escala y  $\sum \sigma_{Xi}^2$  corresponde a la sumatoria de varianzas de los  $i$  ítems. Del mismo modo, otra interpretación posible de este coeficiente es la de un promedio de la covarianza inter-ítems de la escala o promedio de correlación ( $\bar{r}$ ) en el caso de estar estandarizado:

$$\alpha = \frac{k \bar{r}}{1 + (k - 1) \bar{r}} \quad (8)$$

Dado que las puntuaciones en el test ( $X$ ) corresponde a la suma de las puntuaciones en los ítems ( $X_1, X_2, .. X_j$ ), la varianza del test puede expresarse como:

$$\sigma_X^2 = \sigma_{X1}^2 + \sigma_{X2}^2 + \dots + \sigma_{Xj}^2 + 2 \sum Cov(X_i, X_j) \quad (9)$$

Lo que, al desarrollar, permite obtener:

$$\sigma_X^2 = \sum \sigma_{Vj}^2 + \sum \sigma_{ej}^2 + 2 \sum_{i < j} Cov(V_i, V_j) + 2 \sum_{i < j} Cov(e_i, e_j) + 2 \sum_{i < j} Cov(V_i, e_j) \quad (10)$$

Donde  $\sum \sigma_{Vj}^2$  representa la sumatoria de las varianzas verdaderas de los  $j$  ítems,  $\sum \sigma_{ej}^2$  es la sumatoria de las varianzas error de los  $j$  ítems,  $\sum_{i < j} Cov(V_i, V_j)$  corresponde a la covarianza entre la puntuación verdadera de los  $i$  y los  $j$  ítems,  $\sum_{i < j} Cov(e_i, e_j)$  se refiere a la covarianza entre los errores de los ítems y  $\sum_{i \neq j} Cov(V_i, e_j)$  es la covarianza entre la puntuación verdadera y el error. Asumiendo los supuestos de la TCT, tanto las covarianzas entre los errores como la covarianza entre la puntuación verdadera y el error son 0, además asintóticamente el promedio de covarianzas empíricas  $[\sum Cov(X_i, X_j)/(k(k-1))]$  corresponde con el promedio de las covarianzas verdaderas  $[\sum Cov(V_i, V_j)/(k(k-1))]$ ; además considerando que la suma de las varianzas verdaderas de dos variables siempre es igual o mayor que dos por la covarianza entre esas variables  $\sigma_{Vi}^2 + \sigma_{Vj}^2 \geq 2 Cov(V_i, V_j)$  por lo que  $\sigma_{Vi} \sigma_{Vj} \geq Cov(V_i, V_j)$ . En consecuencia, la relación entre la varianza verdadera del test  $\sigma_V^2$  y la varianza total  $\sigma_X^2$  se expresa:

$$\frac{\sigma_V^2}{\sigma_X^2} = \frac{\sum \sigma_{Vj}^2 + \sum Cov(V_i, V_j)}{\sigma_X^2} \geq \frac{\frac{k}{k-1} \sum_{i \neq j} Cov(X_i, X_j)}{\sigma_X^2} = \alpha \quad (11)$$

Entendiéndose así el coeficiente  $\alpha$  como un límite inferior de la fiabilidad. Lo cual solo es cierto bajo determinadas condiciones. A continuación se presentarán los supuestos para que este coeficiente entregue estimaciones adecuadas de la fiabilidad.

## Supuestos del coeficiente alfa

Pese a su difundido uso en investigaciones aplicadas, el coeficiente  $\alpha$  tiene una serie de supuestos que se deben cumplir para garantizar que las estimaciones se correspondan con el verdadero valor de la fiabilidad (Novick & Lewis, 1967), dentro de los más conocidos se encuentran el de Unidimensionalidad, Tau-equivalencia esencial (igual puntuación verdadera para cada ítem), ausencia de correlación entre errores (lo que incluye unidimensionalidad) y normalidad (variables continuas y distribuidas normalmente). Los primeros requisitos han recibido la mayor parte de la atención por los investigadores, sin embargo el impacto de distribuciones no-normales en la estimación de la fiabilidad no ha si tan amplia o profundamente estudiado. A continuación se describirán brevemente los supuestos mencionados, mostrando el efecto que el incumplimiento de cada uno de ellos tiene en la estimación de la fiabilidad del coeficiente  $\alpha$ .

**a) Unidimensionalidad:** De todos los supuestos, el de la Unidimensionalidad resulta de fundamental importancia para la correcta estimación de la fiabilidad (Graham, 2006; Peters, 2014), tanto es así que Lumsden (1976) señala “*The importance of unidimensionality does not depend only on its function as a quasi-mathematical assumption for reliability and scaling theory. The whole conception of psychological testing as measurement depends on it.*” (p. 266). En este contexto la Unidimensionalidad es entendida como la existencia de un único factor común subyacente a los ítems de la escala (un rasgo común que permita explicar las correlaciones o covarianzas entre el conjunto de ítems). En estricto sentido, la Unidimensionalidad perfecta es una condición poco plausible con datos

empíricos (ten Berge & Sočan, 2004), debido a que puede existir covarianza entre los ítems más allá de la atribuible al factor común (e.g., en el caso de correlación entre algún par de errores), existiendo por lo tanto un cierto grado de multidimensionalidad (Sočan, 2000).

Cronbach (1951) pensaba que el coeficiente  $\alpha$  era robusto a la presencia de factores adicionales, siempre y cuando hubiera un factor dominante: “ *$\alpha$  estimates the proportion of the test variance due to all common factors among the items. That is, it reports how much the test score depends upon general and group, rather than item specific, factors*” (p. 320).

La idea de que  $\alpha$  es robusto a la presencia de otros factores de grupo (o específicos) se sustenta en que al aumentar la longitud del test la variabilidad que depende del primer factor se va haciendo cada vez mayor, mientras que la variabilidad que depende de los factores de grupo o específicos va perdiendo peso hasta convertirse en insignificante. De hecho  $\alpha$  se puede definir como la proporción de varianza que explican los factores comunes, en palabras de Cronbach (1951):

*$\alpha$  is the proportion of test variance due to common factors. Our assumption does not hold true when the interitem correlation matrix has rank higher than one. Normally, therefore,  $\alpha$  underestimates the common-factor variance, but not seriously unless the test contains distinct clusters.* (p. 320)

Cronbach (1951) argumenta que lo habitual es encontrar determinada heterogeneidad debida a factores de grupo, aunque no supone en general una mayor amenaza, salvo en estructuras claramente multidimensionales en las que reconoce que  $\alpha$  podría sobre-estimar. Considerando que la interpretación de las puntuaciones es determinada por el primer factor “*The proportion of the test variance due to the first factor among the items is the essential*

*determiner of the interpretability of the scores*” Cronbach recomienda trabajar con datos con un factor dominante “*Interpretability of a test score is enhanced if the score has a high first-factor concentration*” (p. 332).

Cronbach (1951) también analizó el problema de la presencia de ítems con información redundante “*If two items repeat the same information, they are totally redundant*”. Y observó que el uso de ítems redundantes puede derivar en extra-factores, pero mantiene una posición ambigua sobre si éstos deben o no ser incluidos, ya que en algunos casos benefician la medida. La redundancia “*is desirable when accuracy of a single item is low*” (p. 329) si bien también argumenta que “*A set of redundant items can carry much less information than a set of independent items*” (p 330).

En consecuencia, aunque Cronbach minimiza el impacto de la existencia de factores específicos y/o presencia de ítems redundantes, no rechaza el uso del coeficiente  $\alpha$  en estos contextos. Si bien en términos de interpretabilidad de las puntuaciones es mejor que haya un factor dominante, no ve problema en que ítems estén escasamente correlacionados: “*Items with quite low intercorrelations can yield an interpretable scale*” (p. 332)

En este sentido, se considera más realista plantear la posibilidad de unidimensionalidad esencial, evaluando el grado en que el test presenta un factor común (Yang & Green, 2011). Para este fin, el modelo Bifactor, al ser más flexible, permite evaluar el grado en que un test resulta unidimensional al considerar un factor general (común a todos los ítems) junto a factores grupales (Reise, 2012).

Cortina (1993) señala que el coeficiente  $\alpha$  no debe ser utilizado en contextos multidimensionales al observar que si el compuesto de ítems procede de factores ortogonales

(o escasamente correlacionados) el coeficiente  $\alpha$  puede dar valores por encima de .70 cuestionando la interpretación de ese coeficiente global ¿Qué podría significar tales puntuaciones? Así, en situaciones multidimensionales, si lo que interesa es evaluar la fiabilidad atribuida al factor general, el coeficiente  $\alpha$  tiende a generar sobrestimaciones (Zinbarg, Yovel, Revelle, & McDonald, 2006), por el otro lado si se quiere evaluar la proporción de varianza atribuida a todos los factores del modelo dentro de la varianza total del test, el coeficiente  $\alpha$  estaría subestimando esa medida (Revelle & Zinbarg, 2009). Se requiere, por lo tanto, previo a evaluar la fiabilidad, contrastar la estructura subyacente (unidimensional o no) de los datos. Más adelante se presentará el modelo factorial y las estimaciones de fiabilidad que se pueden obtener de él.

**b) Tau-equivalencia esencial:** Se refiere al grado de equivalencia entre los ítems, “Tau” es lo que utiliza Lord y Novick (1968) para identificar a la puntuación verdadera, luego tau-equivalencia es V-equivalencia, en consecuencia se refiere a si las puntuaciones verdaderas son iguales en dos medidas paralelas. Por lo tanto, el modelo tau-equivalente asume que cada ítem de un test mide un mismo constructo con igual puntuación verdadera (o varianza verdadera) pero con la posibilidad de que existan distintas varianzas error; el modelo esencialmente tau-equivalente plantea (esencialmente) lo mismo, pero con la diferencia de que puedan existir constantes (medias) distintas para cada uno los ítems (Graham, 2006; Raykov, 1997). Sin embargo, el que las medias sean distintas o iguales no afecta en la estimación de la fiabilidad (Graham, 2006), por lo que de aquí en adelante nos referiremos a este supuesto sencillamente como tau-equivalencia. En la práctica resulta inusual que este supuesto se cumpla, debido a que al ser altamente restrictivo y poco realista



el ajuste tiende a ser inadecuado con datos empíricos (donde en general las varianzas verdaderas de los ítems presentan diferentes valores). A diferencia del modelo tau-equivalente, el modelo congénérico permite modelar ítems con diferentes varianzas verdaderas (Lord & Novick, 1968), lo que lo hace mucho más flexible y adecuado para datos en psicología. Por lo tanto, resulta habitual trabajar con datos que se ajustan mejor a un modelo congénérico (Teo & Fan, 2013). En la Figura 1 se observa la representación factorial de un modelo unidimensional (a) tau-equivalente con igual puntuación verdadera (valor fijado a .50) y (b) congénérico con puntuación verdadera diferente para cada ítem.

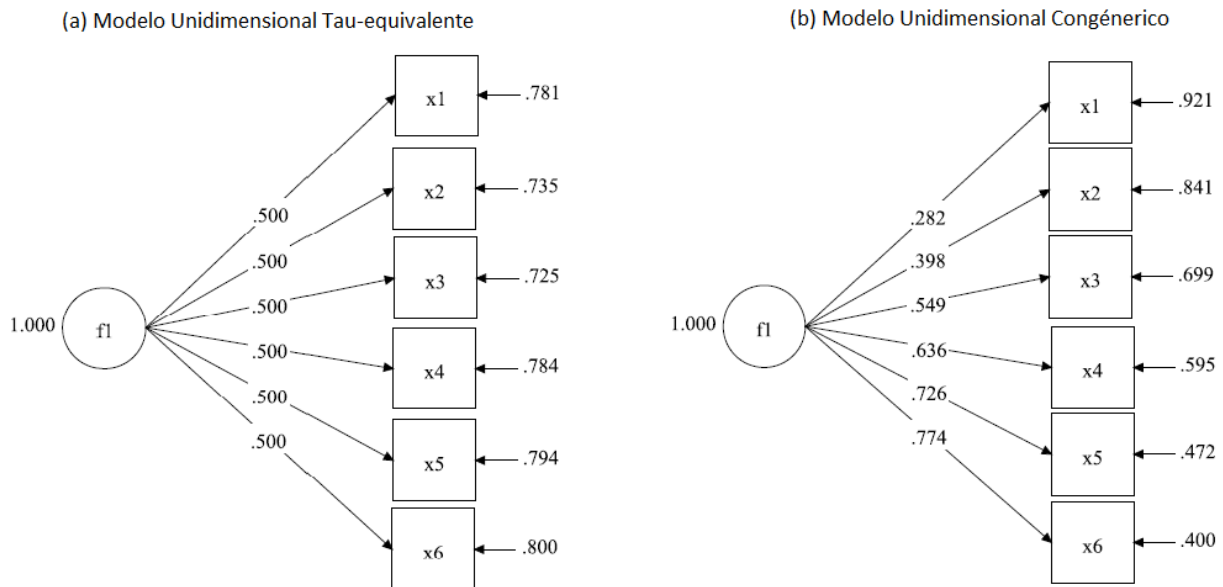


Figura 1: Representación de los modelos (a) Tau-equivalente y (b) Congénéricos con 6 ítems.

Cuando se incumple el supuesto de tau-equivalencia la estimación del coeficiente  $\alpha$  estará negativamente sesgada (Cho & Kim, 2015; Green & Yang, 2009a; Novick & Lewis, 1967; Yang & Green, 2011), siendo más vulnerables aquellos test con un menor número de

ítems (Graham, 2006; Raykov, 1997). Por lo tanto, dependiendo de lo grave que sea la violación de este supuesto, la estimación de la fiabilidad se puede ver negativamente afectada, desde un 0.6% hasta cerca de un 11% (Green & Yang, 2009a). En este caso, ítems con diferente varianza verdadera, el coeficiente  $\alpha$  se convierte en un “lower bound” (límite inferior) de la fiabilidad (Cronbach, 1951).

**c) Ausencia de correlación entre errores:** Desde la TCT los errores de medida de cada ítem no están correlacionados por definición, sin embargo en la práctica resulta poco realista asumir esta hipótesis sin contrastarla, teniendo en cuenta que en general los ítems tienden a compartir, en alguna medida, cierta variabilidad más allá de la atribuible al factor común, ya sea por formato (e.g., tipo Likert, la dirección del ítem, redacción similar, etc.) o por contenido (redundancia, solapamiento, preguntar varias veces lo mismo de diferentes formas, etc.), siendo Streiner (2003) uno de los que nos recuerda que la redundancia de información en los ítems puede generar valores artificialmente inflados de la fiabilidad. Por lo tanto, cuando existen correlaciones positivas entre los términos error de los ítems, el valor del coeficiente  $\alpha$  estará sobre-estimado, mientras que en el caso de que la correlación sea negativa, el coeficiente  $\alpha$  tenderá a infraestimar el valor de la fiabilidad (Raykov, 2001). Como se ha mencionado previamente, la presencia de múltiples correlaciones entre errores dentro de una estructura unidimensional daría cuenta de que esos ítems comparten varianza más allá de la que se puede explicar por el factor común, de este modo y en la medida en que esas correlaciones entre errores permitan modelar una estructura multidimensional más compleja, se pueden realizar estimaciones más precisas de la fiabilidad del factor común, controlando la varianza atribuida a correlaciones entre errores o la varianza atribuida a

factores grupales o específicos. Estos procedimientos corregidos serán presentados más adelante.

Algunos estudios empíricos que incluyen subgrupos de ítems asociados con algún estímulo en particular (como por ejemplo en los denominados *testlet*), han encontrado que el sesgo positivo puede ser pequeño, desde 4 a 7% (Lee & Frisbie, 1999), o más bien grande, llegando a un 18% (Gessaroli & Folske, 2002). Por lo tanto, se debe considerar la correlación entre los elementos error para estimar correctamente la fiabilidad (la corrección será presentada en el apartado de Modelos de Ecuaciones Estructurales). En la Figura 2 se puede observar una representación de un modelo unidimensional con correlación entre errores.

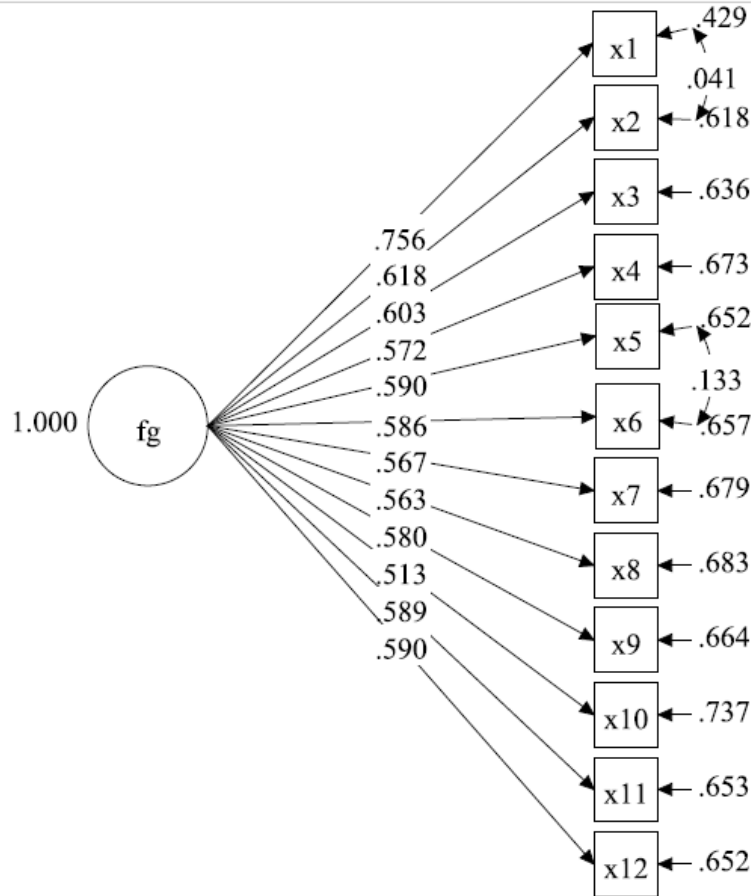


Figura 2: Ejemplo de un modelo unidimensional con correlaciones entre los errores de los ítems 1 y 2 y los ítems 5 y 6.

De este modo, el coeficiente  $\alpha$  puede generar tanto infraestimaciones (violando el supuesto de tau-equivalencia, pero asumiendo el de no correlación entre errores o en presencia de errores negativamente correlacionados) como sobrestimaciones (asumiendo el supuesto de tau-equivalencia pero con correlación positiva entre errores), dependiendo del grado en el que se incumplan estos supuestos combinados (Yang & Green, 2011). Así, las consecuencias que el incumplimiento de estos supuesto tiene en la estimación de la fiabilidad mediante el coeficiente  $\alpha$  no son triviales y se hace evidente la necesidad de evaluar empíricamente el grado en el que los datos cumplen estos supuestos, siendo los Modelos de Ecuaciones Estructurales una herramienta adecuada para ello (Green & Yang, 2009a).

Los tres supuestos mencionados hasta ahora, unidimensionalidad, tau-equivalencia y ausencia de correlación entre errores, se encuentran estrechamente ligados entre sí, por ejemplo, si al evaluar el ajuste de un modelo unidimensional se hace necesario permitir la covariación entre los errores de varios ítems (debido a que existe covarianza sin explicar por el Factor común) cabe la posibilidad de que los datos sean multidimensionales, lo que en la práctica resulta habitual (Zinbarg et al., 2006).

**d) Normalidad:** Este último supuesto que a diferencia de los anteriores es relativamente poco conocido y, por lo tanto, poco estudiado en la literatura referente a la estimación de la fiabilidad (Sheng & Sheng, 2012), asume que los ítems son variables continuas, se encuentran linealmente relacionados y se distribuyen normalmente. En caso de incumplimiento de este supuesto, tanto al trabajar con variables continuas no normales (Sheng & Sheng, 2012) como con variables categóricas (que incumplen además el supuesto de linealidad) asimétricas (Gadermann, Guhn, & Zumbo, 2012; Green & Yang, 2009b; Zumbo, Gadermann, & Zeisser, 2007) se ha mostrado que el coeficiente  $\alpha$  genera infraestimaciones. A pesar de que este supuesto ha sido poco estudiado resulta especialmente relevante debido a que afecta tanto a la correcta estimación del coeficiente  $\alpha$  puntual así como a la posibilidad de establecer intervalos de confianza según el procedimiento clásico (Dunn, Baguley, & Brunsden, 2014). Recientemente Sheng y Sheng (2012) observaron que cuando las distribuciones son asimétricas y/o leptocúrticas se producen sesgos negativos al calcular el coeficiente  $\alpha$ , resultados similares a los presentados por Green & Yang (2009b) al analizar el efectos de distribuciones no normales y diferentes números de categorías en las estimaciones de la fiabilidad. El estudio de los efectos de la asimetría en la estimación de las

evidencias de fiabilidad se hace más relevante cuando se constata que gran parte de las puntuaciones de los instrumentos utilizados en psicología presenta algún grado de asimetría (ver por ejemplo, Ho & Yu, 2014; Micceri, 1989; Norton, Cosco, Doyle, Done, & Sacker, 2013). Pese a ello, este tópico de investigación no ha sido estudiado en profundidad.

Considerando los supuestos anteriormente mencionados, y las consecuencias que se derivan del incumplimiento o no de cada uno de ellos, resulta necesario plantearse la pregunta ¿Existe alguna alternativa más adecuada para la estimación de la fiabilidad que el coeficiente  $\alpha$ ? La respuesta a ese interrogante es que no solo existe una alternativa, por el contrario existen múltiples opciones propuestas para estimar la fiabilidad y superar al clásico  $\alpha$ .

Es posible que el exceso de procedimientos para la estimación de la fiabilidad desarrollados en el último siglo haya oscurecido el debate, contribuyendo también a ello la sencillez de cálculo del coeficiente  $\alpha$ , su disponibilidad en softwares comerciales y la creencia errónea de que podía ser un buen indicador de unidimensionalidad (su valor sería más elevado si la escala era unidimensional), cuando la realidad es justo la contraria como demostró Sitjima (2009a), su aplicación en condiciones multidimensionales supone una violación de la hipótesis de no correlación entre errores que puede conducir a importantes sobre-estimaciones de la fiabilidad (Raykov, 2001) como se ha señalado anteriormente.

A continuación se introducirá al lector a los Modelos de Ecuaciones Estructurales, para contextualizar una de las alternativas que se han propuesto para superar al coeficiente  $\alpha$ , nos referimos al coeficiente  $\omega$  (McDonald, 1999).

## ESTIMACIÓN DE LA FIABILIDAD DESDE LOS MODELOS DE ECUACIONES ESTRUCTURALES

Los Modelos de Ecuaciones Estructurales (*Structural Equation Modeling*, SEM) permiten analizar simultáneamente las relaciones entre múltiples variables dependientes e independientes, tanto observadas como latentes, integrando características de las técnicas de regresión (*path analysis*) y de los modelos usados en psicometría (análisis factorial). Las principales ventajas de trabajar con SEM es que permite plantear un modelo basado en la teoría y evidencia empírica previa que explique los datos obtenidos, de esta forma puede modelar y especificar las relaciones entre dos o más constructos (variables latentes) controlando el error de medida asociado a cada indicador observado, además entrega índices de bondad de ajuste sobre la calidad del modelo evaluado, es decir, sobre la capacidad que tiene el modelo propuesto de reproducir y explicar los datos.

Dentro de SEM, el Análisis Factorial corresponde a un caso particular denominado como el “Modelo de Medida” que incluye las relaciones entre la variable latente y los indicadores (ítems) diseñados para medirla (Brown, 2006). En este contexto, resulta habitual referirse al modelo de medida simplemente como Análisis Factorial Confirmatorio (AFC), que, a diferencia del Análisis Factorial Exploratorio (AFE), permite definir claramente las relaciones entre las variables, contrastando estas relaciones con los datos empíricos y obtener índices de ajuste sobre lo adecuado del modelo. A continuación se expondrá de forma introductoria la conceptualización del *framework* Factorial que replantea las propuestas y ecuaciones de la TCT. Así este modelo se puede expresar formalmente como:

$$X_{ij} = \tau_j + \sum_{k=1}^k \lambda_{jm} F_m + \left( \sqrt{1 - \sum_{k=1}^k \lambda_{jm}^2} \right) e_j \quad (12)$$

donde  $X_{ij}$  es la respuesta de un sujeto  $i$  en el ítem  $j$ ,  $\tau$  representa la constante (intersección, ordenada en el origen o media) del ítem  $j$ ,  $\lambda_{jk}$  es la saturación factorial (parámetro lambda) del ítem  $j$  en el Factor  $m$ , finalmente  $e_j$  es el error aleatorio de medida de cada ítem. Lo anterior, para el caso unidimensional, puede simplificarse como:

$$X_{ij} = \tau_j + \lambda_j F + \sqrt{\psi} e_j \quad (13)$$

En consecuencia, la varianza observada de  $j$  ítem  $\sigma_j^2$  se compone tanto por varianza verdadera (llamada comunalidad)  $\lambda_j^2$  como por varianza error (llamada unicidad)  $1 - \lambda_j^2 = \psi_j$ :

$$\sigma_j^2 = \lambda_j^2 + \psi_j \quad (14)$$

Siguiendo a Raykov (2001), en un modelo congenérico, la fiabilidad de una puntuación compuesta  $X = X_1 + X_2 + \dots + X_k$  es:

$$\rho_X = \frac{\sigma_1(\sum_{i=1}^k \lambda_i)^2}{\left( \sigma_1(\sum_{i=1}^k \lambda_i)^2 + \sum_{i=1}^k \psi_i \right)} \quad (15)$$

Obteniéndose mediante máxima verosimilitud los pesos factoriales o parámetros lambda y las varianzas error psi, al aplicar SEM ajustando los datos a un modelo unidimensional, siendo esta expresión equivalente a la fórmula  $\omega$  de McDonnald.

En caso de existir correlación entre errores para dos elementos, la fórmula se corrige introduciendo dos veces la covarianza (correlación en puntuaciones estandarizadas) entre los



errores en el denominador. La cual es equivalente al coeficiente que se presentará posteriormente  $\omega_b$ :

$$\rho_X = \frac{\sigma_1(\sum_{i=1}^k \lambda_i)^2}{\left(\sigma_1(\sum_{i=1}^k \lambda_i)^2 + \sum_{i=1}^k \psi_i + 2\phi\right)} \quad (16)$$

En el caso de tau-equivalencia (los k ítems con el mismo valor de lambda), siendo la varianza del constructo igual a uno la expresión se reduce a:

$$\rho_X = \frac{(k\lambda)^2}{[(k\lambda)^2 + k\psi + 2\phi]} \quad (17)$$

o bien:

$$\rho_X = \frac{(k\lambda)^2}{[(k\lambda)^2 + k(1 - \lambda^2) + 2\phi]} \quad (18)$$

Estas correcciones a la estimación de la fiabilidad fueron propuestas dentro del marco general de SEM, a continuación se presentan las versiones específicas para el Análisis Factorial.

### **Familia de coeficientes omega**

Por medio de los parámetros lambdas (de discriminación) obtenidos a través del modelo Factorial, se pueden obtener estimaciones adecuadas de la fiabilidad. Uno de los coeficientes que se han propuesto para estimar la fiabilidad dentro de este *framework* es el conocido coeficiente  $\omega$  (McDonald, 1999). De hecho Green y Yang (2009a) señalan que este coeficiente “*blurs the distinctions between reliability and validity*” (p. 132), siendo  $\omega$  la

proporción de varianza de las puntuaciones observadas atribuibles al factor latente. Revelle y Zinbarg (2009) señalan la existencia de al menos dos coeficientes  $\omega$  diferentes, uno llamado  $\omega_t$  (Omega Total) y otro denominado  $\omega_h$  (Omega Jerárquico), además al incorporar las correcciones de Raykov (2001) se presenta un tercer coeficiente  $\omega$  para el caso de errores correlaciones, llamado  $\omega_b$  (Teo & Fan, 2013). El primero de estos coeficientes está basado en la suma de lambdas al cuadrado de todos los factores del modelo (un solo factor en el caso unidimensional), mientras el segundo considera la suma de lambdas al cuadrado del factor general y controla la varianza de factores grupales, por su parte el tercero es conceptualmente similar al primero, pero permite controlar la varianza atribuida a la correlación entre errores. Al fijar la varianza del factor a 1 el coeficiente  $\omega_t$  es definido como:

$$\omega_t = \frac{(\sum \lambda_j)^2}{[(\sum \lambda_j)^2 + \sum (1 - \lambda_j^2)]} = \frac{(\sum \lambda_j)^2}{[(\sum \lambda_j)^2 + (\sum \psi)]} \quad (19)$$

Donde  $\lambda_j$  es la saturación del ítem  $j$ ,  $\lambda_j^2$  es la comunalidad del ítem  $j$  y  $\psi$  equivale a la unicidad. El coeficiente  $\omega_t$  incluyendo los lambdas en su ecuación es aplicable tanto cuando existe tau-equivalencia en los datos (igual saturación factorial para cada ítem, es decir, un único  $\lambda$  para todos los ítems)  $\omega_t$  coincide matemáticamente con el coeficiente  $\alpha$ , y cuando existen ítems con diferente discriminación que son incorporados en la representación del constructo (permite diferentes saturaciones para cada ítem, por lo que estamos en presencia de un modelo congenérico). Por lo tanto,  $\omega_t$  corrige el sesgo de infra-estimación del coeficiente  $\alpha$  cuando se incumple el supuesto de tau-equivalencia. Además diferentes estudios muestran a  $\omega_t$  como una de las mejores alternativas para estimar la fiabilidad (Deng

& Chan, 2016; Revelle & Zinbarg, 2009; Zinbarg, Revelle, Yovel, & Li, 2005; Zinbarg et al., 2006).

Sin embargo cuando existe correlación positiva entre errores, el coeficiente  $\omega_t$  produce una sobre-estimación de la fiabilidad (al igual que  $\alpha$ ). Considerando la corrección propuesta por Raykov (2001) para esta condición, el coeficiente  $\omega_b$  permite incorporar en el denominador ese término de covarianza entre los errores de los ítems:

$$\omega_b = \frac{(\sum \lambda_j)^2}{[(\sum \lambda_j)^2 + (\sum \psi) + 2 (\sum \psi_{ji})]} \quad (20)$$

Donde el término  $\psi_{ji}$  refleja la covariación entre los errores  $j$  e  $i$ . De esta forma se generan estimaciones de la fiabilidad que permiten controlar la existencia de correlación entre los errores de los ítems. Sin embargo, este procedimiento tiene una importante limitación puesto que requiere conocer para qué pares de ítems existe correlación entre errores (covarianza no atribuible al factor general) así como la magnitud y dirección de los mismos. Esto puede ser superado mediante la estimaciones de modelos Bifactor (Green & Yang, 2015; Reise, 2012) en los casos en donde existe una estructura multidimensional subyacente. Bajo este modelo, se puede estimar la fiabilidad del factor general mediante el coeficiente  $\omega_h$  permitiendo corregir la sobrestimación tanto de  $\alpha$  como de  $\omega_t$  en presencia de datos multidimensionales o con errores positivamente correlacionados (Revelle & Zinbarg, 2009; Tarkkonen & Vehkalahti, 2005; Zinbarg et al., 2005). Los coeficientes  $\omega_h$  y  $\omega_t$  son equivalentes en datos estrictamente unidimensionales.

### Estimación de la fiabilidad desde el modelo bifactor

Como se ha mencionado previamente, el modelo Bifactor representa una conceptualización más flexible que los clásicos modelos unidimensionales. Considera tanto un Factor General ( $F_g$ ) común a todos los ítems, como factores grupales ( $F_{s1}, F_{s2}, \dots F_{sk}$ ) que son específicos para ciertos ítems (Jennrich & Bentler, 2011; Reise, 2012), su expresión formal es:

$$X_{ij} = \tau_j + \lambda_{jg} F_g + \lambda_{js1} F_{s1} + \lambda_{js2} F_{s2} + \lambda_{jsk} F_{sk} + \sqrt{\psi} e_j \quad (21)$$

En la Figura 3 se puede observar la representación de este modelo considerando un factor general y tres factores específicos.

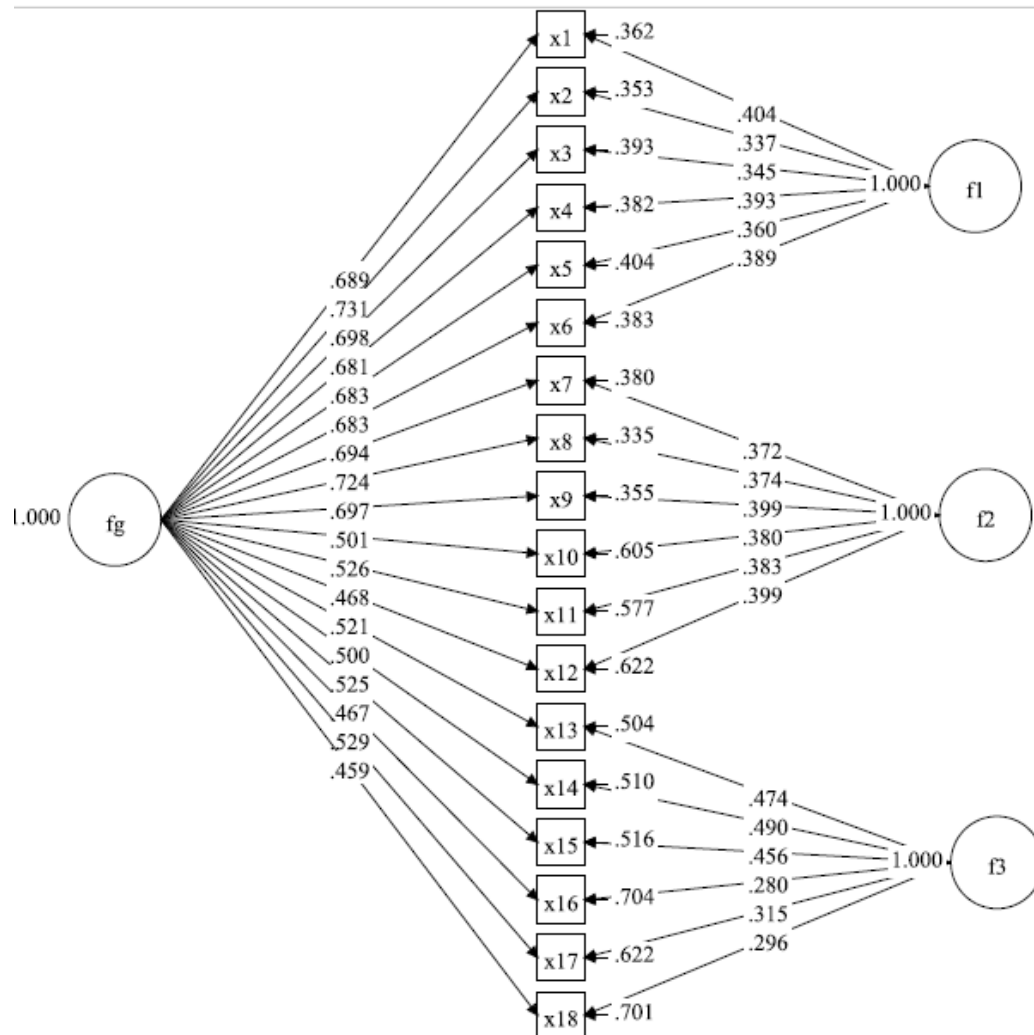


Figura 3: Representación de un Modelo Bifactor considerando 18 ítems y una estructura con un Factor General, común a todos los ítems, y 3 factores grupales cada uno compuesto por 6 ítems.

De acuerdo con Rios & Wells (2014), las ventajas de este modelo, sobre otros multidimensionales, se pueden resumir en las siguientes: a) los factores grupales pueden ser estudiados con independencia del factor general; b) las relaciones entre los factores grupales y sus correspondientes ítems puede ser evaluada; c) se puede evaluar el grado de invarianza tanto en los parámetros de los factores grupales como en el general; y d) es posible evaluar las relaciones entre los factores grupales y criterios externos (variables externas al test) controlando el efecto del factor general y viceversa. Se hace necesario, por lo tanto, evaluar

mediante índices de ajuste el grado en el que un modelo unidimensional o multidimensional, se ajusta a los datos antes de proceder a obtener evidencias de fiabilidad (Crutzen & Peters, 2015). En este contexto, la fiabilidad del factor general se puede estimar con el coeficiente  $\omega_h$  expresado formalmente como:

$$\omega_h = \frac{(\sum \lambda_{jg})^2}{[(\sum \lambda_{jg})^2 + (\sum \lambda_{js1})^2 + (\sum \lambda_{js2})^2 + (\sum \lambda_{jsk})^2 + (\sum \psi)]} \quad (22)$$

La principal ventaja de este coeficiente sobre el coeficiente de Raykov para el caso de correlación entre errores (o el coeficiente  $\omega_b$ ), es que su estimación se realiza sin necesidad de introducir todos los errores uno a uno, ya que surge de la solución bifactorial contrastada. Como se aprecia en su fórmula, el coeficiente  $\omega_h$  permite controlar la varianza atribuible a los factores grupales, entregando una estimación más precisa de la fiabilidad del factor general. Además, al considerar los parámetros lambdas en su ecuación este coeficiente es robusto al incumplimiento de la tau-equivalencia. En un estudio reciente (Green & Yang, 2015) se considera al coeficiente  $\omega_h$  como el estimador más adecuado de la fiabilidad del factor general en este tipo de modelos, además puede ser entendido como el porcentaje de la varianza total del test medido por un único constructo (Revelle & Zinbarg, 2009). Altos valores en este coeficiente indican que los ítems de la escala reflejan unidimensionalidad (esencial), permitiendo interpretar las puntuaciones de la escala de forma global (Reise, 2012) y controlando los problemas de sobre-estimación que tienden a generar  $\alpha$  y  $\omega_t$  en estas situaciones (Zinbarg et al., 2005). Como se ha mencionado previamente, en condiciones unidimensionales los coeficientes  $\omega_h$  y  $\omega_t$  convergen, por lo que se les puede denominar en esos casos sencillamente como  $\omega$ .

En tanto, si lo que interesa al investigador es evaluar la fiabilidad de los factores específicos (o subescalas), controlando el efecto del factor general, el estimador a utilizar es omega subescala  $\omega_s$  el cual se puede expresar como:

$$\omega_s = \frac{(\sum \lambda_{js1})^2}{[(\sum \lambda_{jg})^2 + (\sum \lambda_{js1})^2 + (\sum \psi)]} \quad (23)$$

En el numerador se encuentra la proporción de varianza atribuible al factor específico (en este caso el factor específico 1) y en el denominador se encuentra la varianza total de la escala (descompuesta en varianza del factor general, varianza de los factores específicos y varianza error). Altos valores en este coeficiente indican que el factor específico puede ser interpretado sustantivamente, en tanto bajos valores indican que no se pueden realizar interpretaciones más allá del factor general (Reise, 2012).

Todos estos coeficientes presentan ventajas sobre  $\alpha$ , por ejemplo si lo que se quiere es evaluar la fiabilidad del factor general el coeficiente  $\alpha$  tenderá a generar sobre-estimaciones a diferencia del coeficiente  $\omega_h$  (Zinbarg et al., 2005, 2006), en tanto si el objetivo es evaluar la fiabilidad asociada a todos los factores del modelo ( $\omega_t$ ), el coeficiente  $\alpha$  generará infra-estimaciones (Revelle & Zinbarg, 2009). Considerando que en la práctica es poco probable trabajar con datos perfectamente unidimensionales (Reise, 2012; Sočan, 2000; Yang & Green, 2011) resulta más adecuado evaluar mediante modelos Bifactor el grado de unidimensionalidad que pueda presentar una determinada escala, obteniéndose estimaciones más precisas de la fiabilidad.

## GREATEST LOWER BOUND

Además de los coeficientes mencionados (tanto  $\alpha$  como las diferentes versiones de  $\omega$ ), existen otras alternativas para estimar la fiabilidad, una de ellas es el Greatest Lower Bound (GLB). Sijtsma (2009a) ha mostrado que este coeficiente es uno de los más poderosos estimadores de la fiabilidad. A nivel sustantivo GLB es entendido como la cota (o límite) inferior de la fiabilidad, lo que quiere decir que este coeficiente representa la fiabilidad más pequeña posible dado una matriz  $\mathbf{C}_x$  (Sijtsma, 2009a). Así el verdadero valor de la fiabilidad se encuentra entre GLB y 1, mientras que valores comprendidos entre 0 y GLB no son plausibles (ten Berge & Sočan, 2004). Deducido por Woodhouse and Jackson (1977), desde los supuestos de la TCT ( $\mathbf{C}_x = \mathbf{C}_t + \mathbf{C}_e$ ), parte de que una matriz de covarianzas entre ítems, llamada  $\mathbf{C}_x$ , se descompone en la matriz de la suma de la covarianza verdadera entre los ítems,  $\mathbf{C}_t$ , y en la matriz de covarianza error entre los ítems,  $\mathbf{C}_e$ , esta última matriz es una matriz diagonal, estas tres matrices son semidefinidas positivas, no presentando autovalores negativos (ten Berge & Sočan, 2004), su expresión es:

$$GLB = 1 - \frac{tr(\mathbf{C}_e)}{\sigma_x^2} \quad (24)$$

Donde  $\sigma_x^2$  es la varianza del test y  $tr(\mathbf{C}_e)$  se refiere a la traza de la matriz de varianza error. Esta estimación no tiene una solución analítica, por lo que se ha recurrido a procedimientos factoriales (Sočan, 2000) como Minimum rank factor analysis (procedimiento conocido como glb.fa). Más recientemente se ha elaborado el procedimiento GLB algebraic (GLBa) a partir de un algoritmo elaborado por Andreas Moeltner (Moltner & Revelle, 2015) y que de acuerdo Revelle (2015a) recoge de forma más fiel la definición



original de Jackson & Agunwamba (1977), además tiene como ventaja añadida que permite introducir un vector para ponderar los ítems según su relevancia (Al-Homidan, 2008).

GLB, a pesar de su sólida fundamentación teórica, ha sido poco utilizado en la práctica, si bien algunos estudios empíricos recientes han mostrado un mejor comportamiento de este coeficiente frente a  $\alpha$  (Lila, Oliver, Catalá-Miñana, Galiana, & Gracia, 2014) y frente  $\alpha$  y  $\omega$  (Wilcox, Schoffman, Dowda, & Sharpe, 2014). Sin embargo en muestras menores a 1000 casos, bajo el supuesto de normalidad, se ha observado que este coeficiente tiende a sobre-estimar el verdadero valor de la fiabilidad (Shapiro & ten Berge, 2000); su comportamiento en casos donde se incumple el supuesto de normalidad no ha sido estudiado.

Otros autores (Revelle & Zinbarg, 2009) sostienen que GLB, contrario a su nombre, no es el gran límite inferior de la fiabilidad, existiendo otros coeficientes, como  $\omega_t$  que permiten estimar y comprender la fiabilidad de forma más clara, obteniendo en algunos casos valores superiores a los presentados por GLB. Con respecto a esta crítica, Sijtsma (2009b) contra-argumenta que si GLB con muestras menores a 1000 casos presenta problemas de sobre-estimación, es posible que  $\omega_t$  también sufra de este sesgo, sugiriendo, por lo tanto, realizar estudios que comparen estos estimadores.

## ESTIMACIÓN DE LA FIABILIDAD EN PRESENCIA DE DATOS CATEGÓRICOS

Cuando se utilizan procedimientos factoriales para estimar los parámetros del modelo con datos de estudios empíricos (los cuales generalmente se encuentran categorizados en formato de respuesta dicotómico o politómico) resulta una práctica habitual trabajar con la matriz de correlación producto-momento de Pearson, sin embargo estas correlaciones asumen que las variables son normales, continuas (medidas al menos en intervalo) y linealmente relacionadas; como estos supuestos difícilmente se cumplen en presencia de datos categóricos es posible que la matriz de correlación de Pearson produzca una representación distorsionada de la realidad (Flora & Curran, 2004). Dentro de las razones por las cuales la matriz de correlaciones de Pearson se consideran menos adecuadas para datos categóricos se pueden mencionar las siguientes: asumen que la medición está realizada en escala intervalar, sin embargo en la gran mayoría de los estudios aplicados la escala de los ítems es ordinal, así el incumplimiento de este supuesto implica que la correlación obtenida entre dos variables categorizadas será menor, es decir estará atenuada o negativamente sesgada, a la obtenida por las dos variables si fueran continuas, lo que a su vez afecta a los parámetros  $\lambda$  obtenidos por el modelo factorial, generando infraestimaciones (Asun, Rdz-Navarro, & Alvarado, 2016; Baglin, 2014; Distefano, 2002; Ferrando & Lorenzo-seva, 2013). Por ejemplo, si el parámetro  $\lambda$  es de .75, el valor estimado al trabajar con correlaciones de Pearson puede decaer hasta .65 o incluso obtener valores menores (con mayor sesgo negativo) en las condiciones más extremas (e. g., ítems con asimetría moderada o alta, con pocas categorías de respuesta). Otra posible consecuencia de utilizar la matriz de Pearson para trabajar con ítems dicotómicos es que puede generar “extra Factores” (Reise, Waller, & Comrey, 2000) los cuales agrupan a los ítems por el valor de dificultad que presentan (los

factores por lo tanto reflejan lo fácil o difícil que pueden ser los ítems) pudiendo distorsionar la estructura del cuestionario (Kubinger, 2003). Resulta, por lo tanto, fundamental considerar la naturaleza dicotómica o politómica de los indicadores utilizados en los instrumentos de medición, evitando así el sesgo negativo que produce Pearson al trabajar con variables con pocas categorías de respuesta y en presencia de asimetría (Flora, LaBrish, & Chalmers, 2012; Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2010). Para estas situaciones se ha desarrollado el Análisis Factorial de Ítems (AFI) el cual trabaja sobre la matriz de correlación tetracórica o policórica, dependiendo si el ítem se encuentra en formato dicotómico o politómico respectivamente. Una diferencia importante entre el modelo Factorial clásico y AFI es que el primero asume que la relación entre el factor general (en el caso unidimensional) y los ítem es lineal asumiendo varianzas constantes, por su parte el AFI asume que la relación entre ítems y factor corresponde a una ojiva normal (mediante una función monótona creciente) disminuyendo la varianza a medida que se acerca a los extremos (Ferrando & Lorenzo-Seva, 2014).

Las correlaciones tetracóricas, propuestas por Pearson (1900), son una medida de asociación bivariada, entendida como la correlación producto momento de dos variables continuas no observadas (latentes o ficticias) las cuales se han categorizado ofreciendo solo dos opciones de respuesta en datos observados. Este formato de respuesta en los reactivos se pueden encontrar ampliamente difundidos en los instrumentos de medición en ciencias sociales, un ejemplo sería un ítem que evalúa la actitud de un sujeto hacia determinado tema ofreciendo dos alternativas (a favor o en contra; de acuerdo, en desacuerdo, etc.); este ítem posteriormente puede ser ampliado a 5 categorías (totalmente a favor, parcialmente a favor, ni a favor ni en contra, parcialmente en contra, totalmente en contra); por lo que en este

último caso con un ítem politómico se trabajaría con la matriz policórica (que es una extensión de la tetracórica para más de dos categorías de respuesta). A continuación se describirá brevemente el procedimiento de obtención de este tipo de matrices, el cual es conocido como un proceso de tres etapas (Muthén, 1984): Primero, se considera que a cada ítem observado categórico ( $X_1$ ) se encuentra subyacente una variable continua ficticia y normal ( $X_1^*$ ), la que correspondería a los sujetos si no existiese la limitación impuesta por el formato categórico del ítem; esta variable subyacente es estimada utilizando los valores del o los parámetros umbrales ( $\tau$ ) mediante el método de máxima verosimilitud. Un ejemplo de categorización para el caso dicotómico:

$$\text{Caso Dicotómico: } X_1 = \begin{cases} 0, & X_1^* \leq \tau_1 \\ 1, & X_1^* > \tau_1 \end{cases} \quad (25)$$

Lo anterior puede ser ampliado al caso de ítem politómicos con  $K$  categorías de respuesta:

$$\text{Caso Politómico: } X_1 = \begin{cases} 0, & X_1^* \leq \tau_1 \\ k, & \tau_k < X_1^* \leq \tau_{k+1} \\ K-1, & X_1^* > \tau_K \end{cases} \quad (26)$$

Donde  $\tau_0 = -\infty < \tau_1 < \tau_k < \dots < \tau_{K-1} < \tau_K = \infty$ , estimándose tantos parámetros  $\tau$  como  $K - 1$ . Así para el caso de 5 categorías de respuesta se estimarán 4 umbrales que separarán la distribución continua de la variable subyacente  $X^*$  en 5 espacios.

En una siguiente etapa, con los valores de los umbrales y la distribución bivariada entre cada par de variables se pueden obtener estimaciones, mediante máxima verosimilitud condicional, de las correlaciones tetracóricas (para el ejemplo dicotómico) o policóricas (si son más de dos categorías de respuesta). Finalmente, con la matriz de correlación obtenida se estiman los parámetros del modelo factorial mediante algún método como ULS (Mínimos Cuadrados no Ponderados) o WLSMV (Mínimos Cuadrados Ponderados con media y varianza ajustada). Los procedimientos necesarios para obtener esta matriz de un conjunto de datos y los métodos de estimación se encuentran implementados en diversos softwares de análisis psicométrico tales como LISREL (Jöreskog & Sorbom, 1996) , Mplus (Muthén & Muthén, 1999-2010) , FACTOR (Lorenzo-Seva & Ferrando, 2006, 2013), entre otros.

Dentro de las limitaciones de esta aproximación se encuentran las siguientes: Algunas veces la estimación de la matriz de correlación tetracórica o policórica puede no converger generando matrices no definidas positivamente lo que dificulta el tratamiento matemático de las mismas (Jöreskog, 1994), además las estimaciones generadas pueden presentar alta variabilidad con desviaciones típicas muy altas. Esto dependerá, entre otras cosas, del tamaño de la muestra utilizada, ya que al estimarse estas correlaciones se utilizan tablas de contingencia, lo que provoca que en casos con alto número de categorías (y/o pocos sujetos) algunas casillas no presentan casos suficientes. Así, bajo ciertas condiciones es posible que la aproximación lineal produzca mejores resultados que AFI (Ferrando & Lorenzo-seva, 2013). De cualquier forma, en caso de duda, parece sensato presentar y discutir los resultados obtenidos mediante ambos enfoques (Ferrando & Lorenzo-Seva, 2014). A continuación se presentan las versiones ordinales de los coeficientes de fiabilidad  $\alpha$  y  $\omega$

## Coeficiente alfa ordinal

El coeficiente  $\alpha$  ordinal fue propuesto originalmente por Zumbo, Gadermann y Zeisser (2007) como una alternativa a la versión clásica de  $\alpha$ , la principal diferencia entre ambos es que mientras el último se obtiene de la matriz de correlación producto-momento de Pearson, el primero lo hace sobre matrices de correlación tetracórica o policóricas. Así, el coeficiente  $\alpha$  ordinal es estimado utilizando matrices de correlación policóricas (o tetracóricas si los ítems son dicotómicos). De este modo, la fiabilidad considera la naturaleza ordinal de las escalas tipo Likert (Zumbo et al., 2007).

En diversos artículos, tanto con datos simulados como reales, el equipo de Zumbo (Elosua & Zumbo, 2008; Gadermann et al., 2012; Zumbo et al., 2007) ha mostrado que el coeficiente  $\alpha$  clásico, bajo condiciones de asimetría 0 y -2 y número de opciones de respuesta que varían desde dos a siete, presenta sesgos negativos en la estimación de la fiabilidad, mientras que su versión ordinal presenta una mejor recuperación de la fiabilidad simulada. Otros estudios también han mostrado como el coeficiente  $\alpha$  clásico presenta un mayor sesgo negativo a medida que el número de categorías decrece (Lozano, García-Cueto, & Muñiz, 2008; Maydeu-Olivares, Kramp, García-Forero, Gallardo-Pujol, & Coffman, 2009), estabilizándose en torno a cinco o siete. Del mismo modo, se ha observado que el coeficiente  $\alpha$  es también sensible a distribuciones no normales y en presencia de categorías ordinales de respuesta, produciendo estimaciones negativamente sesgadas (Green & Yang, 2009b; Maydeu-Olivares, Coffman, & Hartmann, 2007). Además, al igual que la versión clásica de  $\alpha$ , el coeficiente  $\alpha$  ordinal mantiene los demás supuestos, por lo que, por ejemplo, en presencia de un modelo congénico el coeficiente ordinal mantendrá un sesgo negativo.

### **Coefficiente omega ordinal**

Con el fin de superar las limitaciones de la versión ordinal del coeficiente  $\alpha$ , se hace necesario estimar la fiabilidad mediante opciones más adecuadas, una opción es el coeficiente  $\omega$  ( $\omega$  total, o cualquiera de sus versiones tanto para modelos con correlación entre errores, como para modelos con estructuras multidimensionales complejas tales como el Bifactor) que, al trabajar sobre la matriz de  $\lambda$  ( $\Lambda$ ) obtenida a partir de la matriz de correlaciones tetracórica o policórica (Elosua & Zumbo, 2008), asume el nivel ordinal en los datos (mientras que las versiones clásicas trabajan generalmente sobre la matriz de Pearson). De este modo se evita el sesgo negativo que produce  $\alpha$  en presencia de saturaciones factoriales con distinto valor (modelo congénico). Corresponde aclarar que si bien el coeficiente  $\omega$  clásico ha sido entendido como la correlación al cuadrado entre la puntuación de la escala y el nivel en el factor (McDonald, 1999), en el caso de su versión ordinal, esa correlación al cuadrado correspondería a las variables ficticias continuas  $X_j^*$  y no al indicador categorizado manifiesto (ítems o reactivos  $X_j$ ).

## OBJETIVOS DE INVESTIGACIÓN

Considerando la abundante literatura sobre las limitaciones del coeficiente  $\alpha$  expuesta en la Introducción, y las opciones que se han presentado para superarlo ( $\omega$ , GLB, GLBa) a continuación se plantean los siguientes objetivos.

- 1- Evaluar el sesgo de  $\alpha$  y de las alternativas  $\omega$  y GLB en condiciones de datos congénéricos y asimétricos, bajo el supuesto de unidimensionalidad
- 2- Determinar el sesgo de  $\alpha$  y de las alternativas  $\omega$  y GLB en presencia de correlaciones positivas entre errores y multidimensionalidad en modelos Bifactor
- 3- Explorar en datos categóricos empíricos el comportamiento de  $\alpha$ ,  $\omega$  y GLB en condiciones ordinales, congénéricas y asimétricas.



## ESTUDIO 1: ESTIMACIONES DE LA FIABILIDAD EN MODELOS UNIDIMENSIONALES

### ESTUDIO 1.1: ESTIMACIONES DE LA FIABILIDAD CON DATOS CONTINUOS<sup>1</sup>

#### Método

##### *Generación de datos*

Los datos fueron generados usando los programas R 3.2 (R Development Core Team, 2013) y RStudio 0.98 (Racine, 2012), de acuerdo al siguiente modelo factorial:

$$X_{ij} = \sum_{k=1}^k \lambda_{jk} * F_k + \sqrt{(1 - \sum_{k=1}^k \lambda_{jk}^2)} * e_j \quad (27)$$

Donde  $X_{ij}$  es la respuesta simulada del sujeto  $i$  en el ítem  $j$ ,  $\lambda_{jk}$  es la saturación factorial del ítem  $j$  en el Factor  $k$  (el cual fue generado mediante el modelo unifactorial),  $F_k$  corresponde al factor latente generado mediante una distribución normal estandarizada (media 0 y varianza 1), y  $e_j$  es el error de medida aleatorio de cada ítem siguiendo igualmente una distribución normal estandarizada.

Ítems asimétricos: Las respuestas simuladas  $X_{ij}$  fueron transformadas para generar distribuciones no-normales usando el procedimiento propuesto por Headrick (2002) aplicando transformaciones polinomiales de quinto orden:

$$Y_{ij} = c_0 + c_1 X_{ij} + c_2 X_{ij}^2 + c_3 X_{ij}^3 + c_4 X_{ij}^4 + c_5 X_{ij}^5 \quad (28)$$

---

<sup>1</sup> El primer estudio de esta tesis ha dado como resultado la publicación de Trizano-Hermosilla & Alvarado (2016) en la revista *Frontiers in Psychology*. A continuación se presentará una reproducción parcial y traducida al español de los resultados en este primer estudio.

Los coeficientes fueron tomados de Sheng y Sheng (2012) e implementados con el objetivo de obtener ítems centrados y asimétricos (Asimetría  $\approx 1$ ), sus valores son:  $c_0 = -0.446924$ ,  $c_1 = 1.242521$ ,  $c_2 = 0.500764$ ,  $c_3 = -0.184710$ ,  $c_4 = -0.017947$ ,  $c_5 = 0.003159$ .

### *Condiciones simuladas*

Para evaluar el desempeño de los coeficientes de fiabilidad ( $\alpha$ ,  $\omega$ , GLB and GLBa) se trabajará con tres tamaños muestrales (250, 500, 1000), dos tamaños de test: corto (6 ítems) y largo (12 ítems), dos condiciones de tau-equivalencia (modelo tau-equivalente y un modelo congénérico) y la incorporación progresiva de ítems asimétricos (desde todos los ítems normales, hasta todos los ítems asimétricos, modificando un ítem cada vez). En el caso del test corto se fijó la fiabilidad en .7308, que en la condición de tau-equivalencia se logra con seis ítems con parámetros de discriminación lambda (saturaciones factoriales) =.558, mientras que en la de datos congénéricos se obtiene fijando los lambda en los valores .3, .4, .5, .6, .7 y .8. En el test largo de 12 ítems la fiabilidad se fijó en .8445 tomando los mismos valores del test corto tanto para la condición de modelo tau-equivalente como para el congénérico (en este caso hubo dos ítems para cada valor de lambda). De este modo, 120 condiciones fueron simuladas con 1000 réplicas en cada caso.

### *Análisis de datos*

Los principales análisis fueron realizados mediante los paquete *Psych* (Revelle, 2015b) y *GPArotation* (Bernaards & Jennrich, 2015), estos paquetes permiten estimar  $\alpha$  y  $\omega$ , además, para la estimación de GLB se utilizó dos aproximaciones computacionales: *glb.fa* (Revelle, 2015a) y *glb.algebraic* (Moltner & Revelle, 2015), este último algoritmo ya ha sido trabajado por autores como Hunt & Bentler (2015).

Con el fin de evaluar la precisión de los distintos estimadores en la recuperación de la fiabilidad, se calculó la raíz media del error al cuadrado (Root Mean Square Error, *RMSE*) y el sesgo (%*bias*). El primero es entendido como el promedio de las diferencias entre la fiabilidad estimada y la simulada y es formalizado como:

$$RMSE = \sqrt{\frac{\sum(\hat{\rho} - \rho)^2}{Nr}} \quad (29)$$

Donde  $\hat{\rho}$  es la fiabilidad estimada por cada coeficiente,  $\rho$  corresponde a la fiabilidad simulada y  $Nr$  al número de réplicas. Se define al % bias como la diferencia entre la media de la fiabilidad estimada (por cada uno de los coeficientes) y la fiabilidad simulada, se define como:

$$\% \text{ bias} = \frac{\sum(\hat{\rho} - \rho)}{Nr} \times 100 \quad (30)$$

En ambos índices, cuanto mayor es el valor, mayor es la imprecisión del estimador, pero a diferencia de RMSE, el sesgo puede ser positivo o negativo, en este caso se obtendría información adicional respecto a si el coeficiente está infra-estimando o sobre-estimando el parámetro de fiabilidad simulado. Siguiendo las recomendaciones de Hoogland y Boomsma (1998) los valores de  $RMSE < 0.05$  y  $\% \text{ bias} < 5\%$  muestran buen ajuste.

## Resultados

Los principales resultados se observan en la Tabla 1 (6 ítems) y Tabla 2 (12 ítems), mostrando los valores de RMSE y Sesgo (% bias) de los coeficientes en condiciones de tau-equivalencia

y datos congenéricos así como el incremento de la asimetría de la distribución del test mediante la incorporación gradual de ítems asimétricos.

Tabla 1: RMSE y Sesgo para las condiciones de tau-equivalencia y datos congenéricos para 6 ítems, tres tamaños muestrales y diferente número de ítems asimétricos

Ítems asimétricos	SK	n	cond	$\alpha$	$\omega$	RMSE		% bias				
						GLB	GLBa	$\alpha$	$\omega$	GLB	GLBa	
Normalidad												
0 ítems	0.0	250	tau	0.03	0.03	<b>0.07</b>	0.04	-0.20	0.00	<b>6.30</b>	3.50	
			cong	0.03	0.02	<b>0.07</b>	0.04	-1.60	-0.10	<b>6.10</b>	3.40	
		500	tau	0.02	0.02	<b>0.06</b>	0.03	0.00	0.00	<b>5.70</b>	2.50	
			cong	0.02	0.02	<b>0.06</b>	0.03	-1.50	0.00	<b>5.50</b>	2.50	
		1000	tau	0.01	0.01	<b>0.06</b>	0.02	-0.10	-0.10	4.90	1.70	
			cong	0.02	0.01	<b>0.06</b>	0.02	-1.50	-0.10	<b>5.00</b>	1.70	
2 ítems	0.3	250	tau	<b>0.06</b>	<b>0.06</b>	0.04	0.03	<b>-5.50</b>	<b>-5.00</b>	2.70	-0.80	
			cong	<b>0.06</b>	0.04	<b>0.05</b>	0.03	<b>-5.40</b>	-3.00	3.90	0.90	
		500	tau	<b>0.06</b>	<b>0.05</b>	0.04	0.03	<b>-5.20</b>	-4.90	1.80	-1.90	
			cong	<b>0.06</b>	0.03	0.04	0.02	<b>-5.10</b>	-2.80	3.20	-0.10	
		1000	tau	<b>0.06</b>	<b>0.05</b>	0.03	0.03	<b>-5.20</b>	<b>-5.00</b>	1.00	-2.70	
			cong	<b>0.05</b>	0.03	0.04	0.02	<b>-5.20</b>	-2.90	2.40	-1.00	
4 ítems	0.6	250	tau	<b>0.11</b>	<b>0.10</b>	0.04	<b>0.06</b>	<b>-10.00</b>	<b>-9.60</b>	-0.80	-4.70	
			cong	<b>0.11</b>	<b>0.09</b>	0.04	<b>0.05</b>	<b>-10.70</b>	<b>-8.10</b>	0.20	-3.50	
		500	tau	<b>0.10</b>	<b>0.10</b>	0.04	<b>0.06</b>	<b>-9.70</b>	<b>-9.40</b>	-1.60	<b>-5.80</b>	
			cong	<b>0.11</b>	<b>0.08</b>	0.03	<b>0.05</b>	<b>-10.50</b>	<b>-8.00</b>	-0.30	-4.60	
		1000	tau	<b>0.10</b>	<b>0.10</b>	0.04	<b>0.07</b>	<b>-9.60</b>	<b>-9.50</b>	-2.50	<b>-6.80</b>	
			cong	<b>0.08</b>	<b>0.05</b>	0.03	0.04	<b>-7.70</b>	<b>-5.20</b>	0.80	-3.10	
6 ítems	0.9	250	tau	<b>0.14</b>	<b>0.14</b>	<b>0.06</b>	<b>0.09</b>	<b>-13.10</b>	<b>-12.80</b>	-3.00	<b>-7.50</b>	
			cong	<b>0.15</b>	<b>0.13</b>	<b>0.05</b>	<b>0.08</b>	<b>-13.80</b>	<b>-11.80</b>	-2.50	<b>-6.80</b>	
		500	tau	<b>0.13</b>	<b>0.13</b>	<b>0.06</b>	<b>0.09</b>	<b>-12.70</b>	<b>-12.50</b>	-4.10	<b>-8.80</b>	
			cong	<b>0.14</b>	<b>0.12</b>	<b>0.05</b>	<b>0.09</b>	<b>-13.50</b>	<b>-11.70</b>	-3.40	<b>-8.00</b>	
		1000	tau	<b>0.13</b>	<b>0.13</b>	<b>0.06</b>	<b>0.10</b>	<b>-12.70</b>	<b>-12.60</b>	<b>-5.00</b>	<b>-9.90</b>	
			cong	<b>0.14</b>	<b>0.12</b>	<b>0.06</b>	<b>0.09</b>	<b>-13.40</b>	<b>-11.60</b>	-4.10	<b>-9.00</b>	

Nota: RMSE = Root Mean Square of Error, SK = asimetría del Test, n = tamaño muestral, cond = Condición, tau = modelo tauequivalente, cong = modelo Congenérico,  $\alpha$  = coeficiente Alfa,  $\omega$  = coeficiente omega, GLB = Greatest Lower Bound (GLB.fa), GLBa = Greatest Lower Bound (GLB.algebraic), en negrita RMSE  $\geq .05$  o % sesgo  $\geq |5\%|$

Tabla 2: RMSE y Sesgo para las condiciones de tau-equivalencia y datos congénéricos para 6 ítems, tres tamaños muestrales y diferente número de ítems asimétricos

				RMSE				% bias			
Ítems asimétricos	SK	n	cond	$\alpha$	$\omega$	GLB	GLBa	$\alpha$	$\omega$	GLB	GLBa
Normalidad											
0 ítems	0.0	250	tau	0.02	0.02	<b>0.05</b>	0.04	-0.10	-0.10	4.80	3.90
			cong	0.02	0.01	<b>0.05</b>	0.04	-0.90	-0.10	4.40	3.80
		500	tau	0.01	0.01	0.04	0.03	-0.10	-0.10	4.20	2.80
			cong	0.01	0.01	0.04	0.03	-0.80	-0.10	3.70	2.70
		1000	tau	0.01	0.01	0.04	0.02	0.00	0.00	3.70	2.00
			cong	0.01	0.01	0.03	0.02	-0.80	0.00	3.10	2.00
2 ítems	0.2	250	tau	0.03	0.02	0.04	0.03	-1.90	-1.70	3.60	2.60
			cong	0.03	0.02	0.04	0.03	-2.00	-1.00	3.70	3.10
		500	tau	0.02	0.02	0.03	0.02	-1.80	-1.70	2.90	1.40
			cong	0.02	0.01	0.03	0.02	-2.00	-1.00	3.00	1.90
		1000	tau	0.02	0.02	0.03	0.01	-1.80	-1.70	2.40	0.60
			cong	0.02	0.01	0.03	0.01	-1.90	-1.00	2.30	1.10
4 ítems	0.3	250	tau	0.04	0.04	0.03	0.02	-3.60	-3.50	2.40	1.40
			cong	0.04	0.03	0.03	0.02	-3.90	-2.80	2.50	1.80
		500	tau	0.04	0.04	0.02	0.01	-3.50	-3.40	1.70	0.10
			cong	0.04	0.03	0.02	0.01	-3.80	-2.70	1.70	0.60
		1000	tau	0.04	0.04	0.02	0.01	-3.50	-3.40	1.00	-0.80
			cong	0.04	0.03	0.02	0.01	-3.80	-2.70	1.00	-0.30
6 ítems	0.5	250	tau	<b>0.06</b>	<b>0.06</b>	0.02	0.02	<b>-5.40</b>	<b>-5.20</b>	1.30	0.20
			cong	<b>0.06</b>	<b>0.06</b>	0.02	0.02	<b>-6.10</b>	<b>-5.20</b>	1.00	0.30
		500	tau	<b>0.05</b>	<b>0.05</b>	0.02	0.02	<b>-5.20</b>	<b>-5.10</b>	0.40	-1.20
			cong	<b>0.06</b>	<b>0.05</b>	0.02	0.02	<b>-6.00</b>	<b>-5.10</b>	-0.10	-1.10
		1000	tau	<b>0.05</b>	<b>0.05</b>	0.02	0.02	<b>-5.20</b>	<b>-5.10</b>	-0.40	-2.20
			cong	<b>0.06</b>	<b>0.05</b>	0.02	0.02	<b>-5.90</b>	<b>-5.10</b>	-0.80	-2.10
8 ítems	0.6	250	tau	<b>0.07</b>	<b>0.07</b>	0.02	0.02	<b>-6.90</b>	<b>-6.70</b>	0.20	-1.00
			cong	<b>0.08</b>	<b>0.07</b>	0.02	0.02	<b>-7.30</b>	<b>-6.20</b>	0.20	-0.50
		500	tau	<b>0.07</b>	<b>0.07</b>	0.02	0.03	<b>-6.80</b>	<b>-6.70</b>	-0.70	-2.50
			cong	<b>0.07</b>	<b>0.06</b>	0.02	0.02	<b>-7.20</b>	<b>-6.10</b>	-0.80	-1.90
		1000	tau	<b>0.07</b>	<b>0.07</b>	0.02	0.04	<b>-6.70</b>	<b>-6.60</b>	-1.50	-3.60
			cong	<b>0.07</b>	<b>0.06</b>	0.02	0.03	<b>-7.10</b>	<b>-6.10</b>	-1.70	-3.00
10 ítems	0.8	250	tau	<b>0.09</b>	<b>0.09</b>	0.03	0.03	<b>-8.30</b>	<b>-8.10</b>	-0.70	-2.00
			cong	<b>0.09</b>	<b>0.08</b>	0.02	0.03	<b>-8.80</b>	<b>-7.70</b>	-0.90	-1.60
		500	tau	<b>0.08</b>	<b>0.08</b>	0.03	0.04	<b>-8.10</b>	<b>-8.00</b>	-1.70	-3.60
			cong	<b>0.09</b>	<b>0.08</b>	0.03	0.04	<b>-8.60</b>	<b>-7.60</b>	-1.60	-3.10
		1000	tau	<b>0.08</b>	<b>0.08</b>	0.03	<b>0.05</b>	<b>-8.10</b>	<b>-8.00</b>	-2.40	-4.80
			cong	<b>0.09</b>	<b>0.08</b>	0.03	0.04	<b>-8.60</b>	<b>-7.60</b>	-2.40	-4.30
12 ítems	1.0	250	tau	<b>0.10</b>	<b>0.10</b>	0.03	0.04	<b>-9.40</b>	<b>-9.20</b>	-1.40	-2.80
			cong	<b>0.10</b>	<b>0.09</b>	0.03	0.03	<b>-9.50</b>	<b>-8.50</b>	-1.30	-2.30
		500	tau	<b>0.09</b>	<b>0.09</b>	0.03	<b>0.05</b>	<b>-9.20</b>	<b>-9.10</b>	-2.30	-4.60
			cong	<b>0.10</b>	<b>0.09</b>	0.03	0.04	<b>-9.40</b>	<b>-8.40</b>	-2.30	-3.90
		1000	tau	<b>0.09</b>	<b>0.09</b>	0.04	<b>0.06</b>	<b>-9.20</b>	<b>-9.10</b>	-3.20	<b>-5.90</b>
			cong	<b>0.09</b>	<b>0.08</b>	0.04	<b>0.05</b>	<b>-9.30</b>	<b>-8.30</b>	-3.20	<b>-5.20</b>

Nota: RMSE = Root Mean Square of Error, SK = asimetría del test, n = tamaño muestral, cond = Condición, tau = modelo tauequivalente, cong = modelo Congénérico,  $\alpha$  = coeficiente Alfa,  $\omega$  = coeficiente omega, GLB = Greatest Lower Bound (GLB.fa), GLBa = Greatest Lower Bound (GLB.algebraic), en negrita RMSE  $\geq .05$  o % sesgo  $\geq |5\%|$

Solo bajo condiciones de tau-equivalencia y normalidad (asimetría < .2) se observa que el coeficiente  $\alpha$  estima la fiabilidad simulada correctamente, al igual que  $\omega$ ; para la condición congénica  $\omega$  corrige la subestimación de  $\alpha$ . Por su parte GLB y GLBa presentan un sesgo positivo bajo normalidad, sin embargo GLBa muestra aproximadamente ½ del % sesgo que GLB (ver Tabla 1). Si consideramos el tamaño de la muestra, observamos que a medida que aumenta, el sesgo positivo de GLB y GLBa disminuye, pero nunca desaparece.

A medida que aumenta la asimetría del test, vemos en la Tabla 1 que tanto  $\alpha$  como  $\omega$  presentan un desempeño inaceptable con un RMSE creciente y subestimaciones que pueden alcanzar un sesgo > 13% para el coeficiente  $\alpha$  (entre 1% y 2% menor para  $\omega$ ). Comparativamente los coeficientes GLB y GLBa presentan un RMSE menor cuando aumenta la asimetría del test mediante la incorporación de ítem asimétricos (ver Tabla 1 y 2). De este modo, GLB presenta mejores estimaciones que  $\alpha$  y  $\omega$  cuando el valor de asimetría del test es de alrededor de .30; Funcionamiento similar a GLBa que presenta buenas estimaciones con un valor de asimetría de prueba alrededor de .20 o .30 Sin embargo, cuando el valor de asimetría aumenta a .50 o .60, GLB presenta un mejor rendimiento que GLBa. El tamaño del test (6 o 12 ítems) tiene un efecto mucho más importante que el tamaño muestral en la precisión de las estimaciones.

## **Conclusión**

Cuando la puntuación total del test está normalmente distribuida (e.g., todos los ítems siguen una distribución normal.)  $\omega$  debería ser la primera opción, seguido de  $\alpha$ , ya que evitan los

problemas de sobrestimación que genera GLB. Sin embargo, cuando existe asimetría baja o moderada en el test (o en los ítems), el coeficiente GLBa parece una adecuada opción. Por su parte, el coeficiente GLB es recomendado cuando existe una alta proporción de ítems asimétricos, ya que bajo esas condiciones las estimaciones producidas por  $\alpha$  y  $\omega$  no son aceptables, independientemente del tamaño muestral.

## ESTUDIO 1.2: ESTIMACIÓN DE LA FIABILIDAD CON DATOS CATEGÓRICOS

### Método

#### *Generación de los datos:*

Los datos fueron generados utilizando el software R 3.2 (R Development Core Team, 2013) y RStudio 0.98 (Racine, 2012), siguiendo el modelo factorial:

$$X_{ij} = \sum_{k=1}^k \lambda_{jk} * F_k + \sqrt{(1 - \sum_{k=1}^k \lambda_{jk}^2)} * e_j \quad (31)$$

Donde  $X_{ij}$  corresponde a la respuesta simulada del sujeto  $i$  en el ítem  $j$ ,  $\lambda_{jk}$  es la saturación factorial del ítem  $j$  en el Factor  $k$  (el cual fue generado mediante el modelo unifactorial),  $F_k$  corresponde al factor latente generado mediante una distribución normal estandarizada (media 0 y varianza 1), y  $e_j$  es el error de medida aleatorio de cada ítem siguiendo igualmente una distribución normal estandarizada.

Posteriormente, las respuestas continuas simuladas de los sujetos fueron recodificadas para generar datos categóricos, con dos, tres, cuatro y cinco categorías. Cada una de estas condiciones fueron creadas mediante puntos de corte en las distribuciones continuas,

construyendo variables categóricas con asimetría cercana a 0 (para cada número de categorías) así como distribuciones categóricas asimétricas (con valores en torno a -1). Un ejemplo de este tipo de ítems asimétricos se puede observar en la Figura 4.

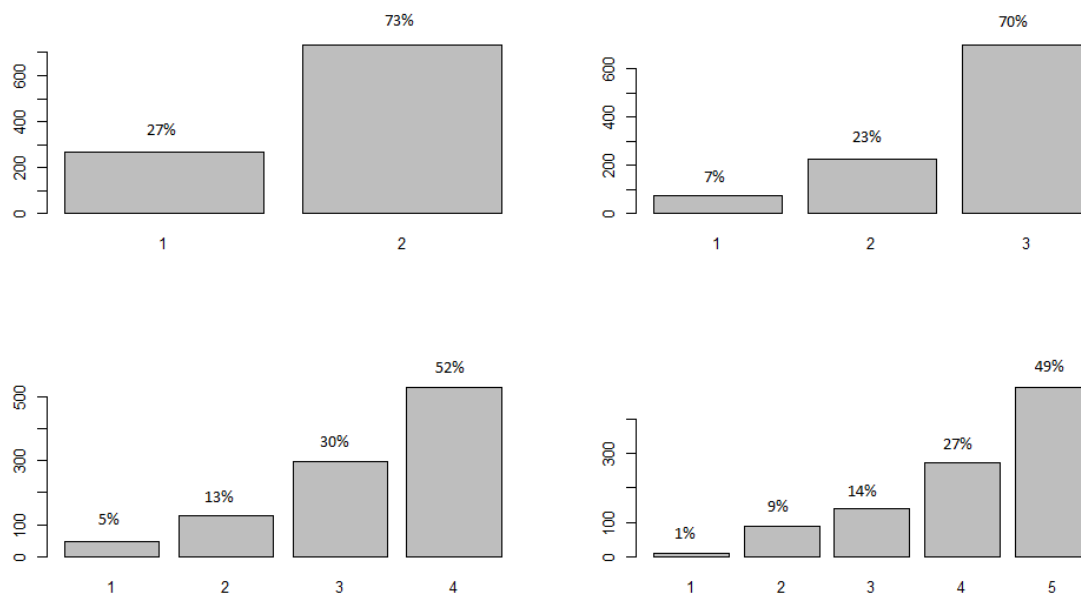


Figura 4: Ejemplos de ítems asimétricos con dos, tres, cuatro y cinco categorías de respuesta.

### *Condiciones simuladas*

Para evaluar el desempeño de los coeficientes de fiabilidad  $\alpha$ ,  $\omega$  (clásicos y ordinales), GLB y GLBa, se trabajará con tres tamaños muestrales (250, 500, 1000), un único tamaño de test de 6 ítems generados bajo un modelo congénico (fijando los lambda en los valores .3, .4, .5, .6, .7 y .8.) y la incorporación progresiva de ítems asimétricos (desde todos los ítems no asimétricos, hasta todos los ítems asimétricos, modificando un ítem cada vez, resultando en 7 condiciones de asimetría en el test). Además, como se ha mencionado antes, se incorporan desde dos hasta cinco categorías. En este caso el valor de la fiabilidad simulada



fue de .7308. En resumen se generaron 84 condiciones: 4 números de categorías, 3 tamaños muestrales y 7 condiciones de asimetría del test. Se realizaron 1000 réplicas.

### *Análisis de datos*

Los principales análisis fueron realizados mediante los paquete *Psych* (Revelle, 2015b) y *GPArotation* (Bernaards & Jennrich, 2015), estos paquetes permiten estimar  $\alpha$  y  $\omega$ , además, para la estimación de GLB se utilizó dos aproximaciones computacionales: *glb.fa* (Revelle, 2015a) y *glb.algebraic* (Moltner & Revelle, 2015), este último algoritmo ya ha sido trabajado por autores como Hunt & Bentler (2015). Para el análisis de los coeficientes ordinales se utilizó el comando “omega” con la opción “*Polychor*” activada.

Con el fin de evaluar la precisión de los distintos estimadores en la recuperación de la fiabilidad, se calculó la raíz media del error al cuadrado (Root Mean Square Error, *RMSE*) y el sesgo. El primero es entendido como el promedio de las diferencias entre la fiabilidad estimada y la simulada y es formalizado como:

$$RMSE = \sqrt{\frac{\sum(\hat{\rho} - \rho)^2}{n}} \quad (32)$$

Donde  $\hat{\rho}$  corresponde a la fiabilidad estimada por cada coeficiente,  $\rho$  a la fiabilidad simulada y  $n$  al número de réplicas. Por su parte, el sesgo se entiende como la diferencia entre la media de la fiabilidad estimada y la fiabilidad simulada y se define como:

$$bias = \frac{\sum(\hat{\rho} - \rho)}{n} \quad (33)$$

En ambos índices, cuanto mayor es el valor mayor es la imprecisión del estimador, pero a diferencia de RMSE, el sesgo puede ser positivo o negativo, en este caso se obtendría

información adicional respecto a si el coeficiente está infra-estimando o sobre-estimando el parámetro de fiabilidad simulado.

## **Resultados**

Los resultados serán presentados por el orden de número de categorías, desde dos hasta cinco.

### *2 categorías de respuesta*

En las Tablas 3 y 4 se observan los valores de asimetría, curtosis y las estimaciones medias obtenidas con cada uno de los coeficientes planteados junto a los resultados del RMSE y el Sesgo, en la condición de ítems con solo dos opciones de respuesta.

En general se aprecia que los valores medios obtenidos tanto por  $\alpha$  como por  $\omega$  son considerablemente más bajos que la fiabilidad simulada, incluso en condiciones donde la asimetría es 0, situación que se va agravando a medida que se van transformando los ítems en asimétricos, para  $\alpha$  la infraestimación de la fiabilidad puede llegar a valores de .171, mientras que para  $\omega$  la infrestimación llega a .156. Se observan valores de RMSE desde .15 hasta .18 y Sesgo desde -.15 hasta -.18, lo que implica que la infraestimación realizada en condiciones binarias es bastante severa. Para  $\omega$  el resultado es bastante similar, con RMSE con valores comprendidos entre .14 y .16 y Segas de -.13 hasta -.16. Para ambos estimadores los valores resultan demasiado altos, por lo que son inaceptables. En tanto para los coeficientes GLB se observa que en la condición de 0 ítems asimétricos presentan valores medios menores que la fiabilidad simulada, sin embargo son sustancialmente mejores que  $\alpha$  o  $\omega$ . A medida que aumenta la asimetría, se observa un efecto negativo en estos coeficientes. Del mismo modo, se aprecia un efecto del tamaño muestral, donde se presentan

estimaciones más altas y cercanas a la fiabilidad simulada, en tamaños muestral de 250 pero decrecen a medida que el tamaño muestral aumenta. Presentan valores de RMSE y Sesgo menores que los de  $\alpha$  y  $\omega$ , resultando mejores los de GLB (RMSE y Sesgo menores a  $|.10|$ ) que los de GLBa (RMSE y Sesgo entre  $|.8|$  y  $|.13|$ ). Finalmente en relación a la versión ordinal de los coeficientes  $\alpha$  y  $\omega$  se observa una recuperación más adecuada del parámetro de fiabilidad, con independencia de la asimetría o del tamaño muestral, mostrando valores de RMSE entre  $.2$  y  $.5$ . Sin embargo, el coeficiente  $\alpha$  ordinal se ve afectado por la constante en este modelo: se está trabajando con datos congenéricos, incumpliendo así uno de los supuestos de este coeficiente (lo que afecta tanto a la versión lineal como a la ordinal). Tal y como se había visto anteriormente, bajo un modelo congénico el coeficiente  $\alpha$  presenta sesgos negativos, infraestimando el parámetro de fiabilidad simulado (Sesgo de  $-.02$ ), no sucediendo lo mismo con  $\omega$  ordinal (sesgo prácticamente de  $0$ ).

Tabla 3: 2 categorías 6 ítems congénico (fiabilidad =  $.731$ )

Ítems asimétricos	n	Asimetría	Curtosis	$\alpha$	$\omega$	GLB	GLBa	$\alpha$ ordinal	$\omega$ ordinal
0	250	0.006	-1.786	.581	.596	.692	.649	.715	.733
	500	0.009	-1.810	.580	.594	.681	.632	.715	.730
	1000	0.007	-1.832	.582	.595	.673	.622	.716	.731
1	250	-0.179	-1.618	.577	.593	.691	.651	.715	.733
	500	-0.173	-1.635	.576	.591	.679	.634	.715	.731
	1000	-0.178	-1.655	.579	.592	.670	.625	.716	.731
2	250	-0.367	-1.443	.570	.587	.686	.649	.714	.732
	500	-0.361	-1.461	.570	.586	.675	.633	.714	.730
	1000	-0.364	-1.477	.573	.588	.667	.624	.716	.731
3	250	-0.553	-1.267	.564	.581	.682	.646	.713	.731
	500	-0.547	-1.285	.564	.580	.670	.630	.713	.730
	1000	-0.550	-1.295	.566	.581	.663	.620	.715	.731
4	250	-0.741	-1.087	.559	.576	.680	.641	.713	.732
	500	-0.733	-1.111	.559	.574	.667	.625	.713	.730
	1000	-0.735	-1.119	.561	.575	.659	.614	.715	.731
5	250	-0.928	-0.908	.555	.571	.676	.633	.712	.732

	500	-0.919	-0.938	.555	.569	.663	.616	.713	.730
	1000	-0.919	-0.941	.557	.570	.654	.605	.715	.730
6	250	-1.113	-0.737	.559	.576	.681	.634	.711	.730
	500	-1.107	-0.762	.560	.575	.666	.617	.713	.730
	1000	-1.108	-0.766	.560	.574	.658	.604	.715	.730

Nota: n = tamaño muestral,  $\alpha$  = Coeficiente Alfa;  $\omega$  = Coeficiente Omega; GLB = Coeficiente GLB ("glb.fa"); GLBa = Coeficiente GLB ("glb.algebraic");  $\omega$  ordinal = Coeficiente Omega Ordinal;  $\alpha$  ordinal = Coeficiente Alfa ordinal

Tabla 4: RMSE y Sesgo de los coeficientes de fiabilidad para la condición de 2 categorías

ítems asimétricos	n	RMSE $\alpha$	RMSE $\omega$	RMSE GLB	RMSE GLBa	RMSE $\alpha$ or	RMSE $\omega$ or
0	250	.16	.14	.06	.09	.04	.03
	500	.15	.14	.07	.10	.03	.02
	1000	.15	.14	.07	.11	.02	.02
1	250	.16	.14	.06	.09	.04	.03
	500	.16	.14	.07	.10	.03	.02
	1000	.15	.14	.07	.11	.02	.02
2	250	.17	.15	.06	.09	.04	.03
	500	.16	.15	.07	.10	.03	.03
	1000	.16	.14	.08	.11	.02	.02
3	250	.17	.15	.07	.09	.04	.03
	500	.17	.15	.08	.10	.03	.03
	1000	.17	.15	.08	.11	.03	.02
4	250	.18	.16	.07	.10	.04	.04
	500	.18	.16	.08	.11	.03	.03
	1000	.17	.16	.08	.12	.02	.02
5	250	.18	.16	.07	.11	.04	.03
	500	.18	.16	.08	.12	.03	.03
	1000	.18	.16	.09	.13	.03	.02
6	250	.18	.16	.07	.11	.05	.04
	500	.17	.16	.08	.12	.03	.03
	1000	.17	.16	.08	.13	.03	.02
Media		.17	.15	.07	.11	.03	.03
	n	Bias $\alpha$	Bias $\omega$	Bias GLB	Bias GLBa	Bias $\alpha$ or	Bias $\omega$ or
0	250	-.15	-.13	-.04	-.08	-.02	.00
	500	-.15	-.14	-.05	-.10	-.02	.00
	1000	-.15	-.14	-.06	-.11	-.01	.00
1	250	-.15	-.14	-.04	-.08	-.02	.00
	500	-.15	-.14	-.05	-.10	-.02	.00
	1000	-.15	-.14	-.06	-.11	-.01	.00
2	250	-.16	-.14	-.05	-.08	-.02	.00
	500	-.16	-.15	-.06	-.10	-.02	.00
	1000	-.16	-.14	-.06	-.11	-.01	.00
3	250	-.17	-.15	-.05	-.09	-.02	.00
	500	-.17	-.15	-.06	-.10	-.02	.00
	1000	-.16	-.15	-.07	-.11	-.02	.00
4	250	-.17	-.16	-.05	-.09	-.02	.00
	500	-.17	-.16	-.06	-.11	-.02	.00

5	1000	<b>-.17</b>	<b>-.16</b>	<b>-.07</b>	<b>-.12</b>	-.02	.00
	250	<b>-.18</b>	<b>-.16</b>	<b>-.06</b>	<b>-.10</b>	-.02	.00
	500	<b>-.18</b>	<b>-.16</b>	<b>-.07</b>	<b>-.12</b>	-.02	.00
	1000	<b>-.17</b>	<b>-.16</b>	<b>-.08</b>	<b>-.13</b>	-.02	.00
6	250	<b>-.17</b>	<b>-.16</b>	<b>-.05</b>	<b>-.10</b>	-.02	.00
	500	<b>-.17</b>	<b>-.16</b>	<b>-.06</b>	<b>-.11</b>	-.02	.00
	1000	<b>-.17</b>	<b>-.16</b>	<b>-.07</b>	<b>-.13</b>	-.02	.00
	Media	-.16	-.15	-.06	-.10	-.02	.00

Nota: n = tamaño muestral, RMSE = raíz media del error al cuadrado, Bias = Sesgo,  $\alpha$  = Coeficiente alfa,  $\omega$  = Coeficiente Omega, GLB = glb.fa, GLBa = glb.algebraic,  $\alpha$  or = Coeficiente alfa ordinal,  $\omega$  or = Coeficiente Omega ordinal; en negrita los valores superiores a  $|\cdot|.05|$

### *3 categorías de respuesta*

Para el caso de tres opciones de respuesta, en la Tabla 5 y en la Tabla 6 se presentan los mismos estadísticos que en el caso anterior. Comparativamente con el caso de dos categorías, las estimaciones de alfa y omega han mejorado en aproximadamente .05. Del mismo modo se ven afectados negativamente por el incremento en el número de ítems asimétricos dentro del total del test (con valor medio de RMSE de .12 para  $\alpha$  y de .11 para  $\omega$ , en tanto que el SESGO llega a valores medios de -.12 para el primero y -.10 para el segundo, manteniéndose en valores inaceptables al igual que en la condición de dos categorías). Por su parte, GLB y GLBa presentan en general un mejor funcionamiento que los dos anteriores coeficientes, siendo superior el desempeño de GLB (con RMSE medio de .047 y Sesgo de -.024, implicando una leve infraestimación en condiciones de asimetría) que el de GLBa (Con RMSE medio de .078 y Sesgo medio de -.073). Se observa que los valores obtenidos de  $\alpha$  y  $\omega$  ordinal son similares al caso de 2 ítems, manteniéndose constante el patrón (leve Sesgo negativo en torno a -.02 para  $\alpha$  ordinal al incumplir el supuesto de tau-equivalencia).

Tabla 5: 3 categorías 6 ítems congénico (fiabilidad = .731)

ítems asimétricos	n	as	cr	$\alpha$	$\omega$	glb	glba	$\alpha$ ordinal	$\omega$ ordinal
0	250	-0.001	-0.244	.633	.648	.730	.692	.717	.734
	500	-0.001	0.045	.626	.639	.715	.671	.718	.734
	1000	-0.002	0.365	.614	.627	.699	.651	.716	.731
1	250	-0.230	-0.073	.628	.644	.729	.687	.717	.733
	500	-0.230	0.172	.621	.635	.713	.664	.718	.733
	1000	-0.232	0.442	.610	.624	.696	.641	.716	.731
2	250	-0.459	0.101	.621	.638	.723	.682	.716	.733
	500	-0.459	0.296	.615	.630	.710	.657	.717	.733
	1000	-0.462	0.517	.606	.620	.693	.633	.716	.731
3	250	-0.688	0.277	.613	.630	.718	.676	.715	.732
	500	-0.689	0.425	.609	.624	.705	.651	.717	.733
	1000	-0.691	0.585	.602	.616	.691	.627	.716	.731
4	250	-0.918	0.451	.606	.623	.716	.671	.715	.732
	500	-0.919	0.551	.603	.618	.700	.648	.716	.732
	1000	-0.922	0.660	.599	.612	.686	.627	.716	.731
5	250	-1.148	0.626	.603	.618	.713	.670	.714	.731
	500	-1.149	0.677	.602	.616	.699	.650	.715	.731
	1000	-1.152	0.731	.599	.612	.686	.633	.715	.730
6	250	-1.376	0.802	.607	.624	.717	.678	.712	.730
	500	-1.377	0.800	.609	.625	.707	.663	.714	.730
	1000	-1.381	0.808	.609	.624	.624	.696	.715	.730

Nota: n = tamaño muestral,  $\alpha$  = Coeficiente Alfa;  $\omega$  = Coeficiente Omega; GLB = Coeficiente GLB ("glb.fa"); GLBa = Coeficiente GLB ("glb.algebraic");  $\omega$  ordinal = Coeficiente Omega Ordinal;  $\alpha$  ordinal = Coeficiente Alfa ordinal

Tabla 6: RMSE y Sesgo de los coeficientes de fiabilidad para la condición de 3 categorías

ítems asimétricos	n	RMSE $\alpha$	RMSE $\omega$	RMSE GLB	RMSE GLBa	RMSE $\alpha$ or	RMSE $\omega$ or
0	250	.104	.089	.039	.050	.035	.030
	500	.108	.095	.041	.064	.027	.022
	1000	.118	.105	.050	.082	.022	.016
1	250	.109	.093	.037	.054	.036	.030
	500	.113	.099	.041	.072	.027	.022
	1000	.122	.109	.051	.092	.023	.016
2	250	.116	.099	.040	.059	.037	.031
	500	.119	.104	.043	.078	.028	.022
	1000	.126	.112	.054	.100	.023	.016
3	250	.124	.106	.042	.065	.038	.031
	500	.125	.109	.046	.084	.028	.022
	1000	.131	.116	.056	.105	.023	.016
4	250	.130	.114	.043	.069	.039	.031
	500	.131	.115	.050	.087	.029	.023

	1000	<b>.134</b>	<b>.120</b>	<b>.059</b>	<b>.106</b>	.023	.016
5	250	<b>.135</b>	<b>.119</b>	.046	<b>.072</b>	.041	.033
	500	<b>.132</b>	<b>.118</b>	.051	<b>.086</b>	.030	.024
	1000	<b>.133</b>	<b>.120</b>	<b>.059</b>	<b>.100</b>	.024	.016
6	250	<b>.131</b>	<b>.114</b>	.046	<b>.065</b>	.043	.034
	500	<b>.126</b>	<b>.110</b>	.046	<b>.074</b>	.032	.024
	1000	<b>.124</b>	<b>.109</b>	<b>.052</b>	<b>.083</b>	.024	.017
	Media	<b>.123</b>	<b>.108</b>	.047	.078	.030	.023

ítems asimétricos	n	Bias $\alpha$	Bias $\omega$	Bias GLB	Bias GLBa	Bias $\alpha$ or	Bias $\omega$ or
0	250	<b>-.098</b>	<b>-.083</b>	-.001	-.039	-.014	.003
	500	<b>-.105</b>	<b>-.092</b>	-.016	<b>-.060</b>	-.013	.003
	1000	<b>-.117</b>	<b>-.104</b>	-.032	<b>-.080</b>	-.015	.000
1	250	<b>-.103</b>	<b>-.087</b>	-.002	-.044	-.014	.003
	500	<b>-.110</b>	<b>-.096</b>	-.018	<b>-.067</b>	-.013	.002
	1000	<b>-.121</b>	<b>-.107</b>	-.035	<b>-.090</b>	-.015	.000
2	250	<b>-.110</b>	<b>-.093</b>	-.008	-.049	-.015	.002
	500	<b>-.116</b>	<b>-.101</b>	-.021	<b>-.074</b>	-.014	.002
	1000	<b>-.125</b>	<b>-.110</b>	-.038	<b>-.098</b>	-.015	.000
3	250	<b>-.118</b>	<b>-.101</b>	-.013	<b>-.055</b>	-.016	.002
	500	<b>-.122</b>	<b>-.106</b>	-.026	<b>-.080</b>	-.014	.002
	1000	<b>-.129</b>	<b>-.115</b>	-.040	<b>-.103</b>	-.015	.000
4	250	<b>-.125</b>	<b>-.108</b>	-.015	<b>-.060</b>	-.016	.001
	500	<b>-.128</b>	<b>-.112</b>	-.031	<b>-.083</b>	-.015	.001
	1000	<b>-.132</b>	<b>-.119</b>	-.045	<b>-.104</b>	-.015	.000
5	250	<b>-.128</b>	<b>-.113</b>	-.018	<b>-.061</b>	-.017	.000
	500	<b>-.129</b>	<b>-.115</b>	-.032	<b>-.081</b>	-.016	.001
	1000	<b>-.131</b>	<b>-.119</b>	-.045	<b>-.098</b>	-.016	-.001
6	250	<b>-.124</b>	<b>-.106</b>	-.014	<b>-.052</b>	-.019	-.001
	500	<b>-.122</b>	<b>-.106</b>	-.024	<b>-.068</b>	-.017	.000
	1000	<b>-.122</b>	<b>-.107</b>	-.035	<b>-.080</b>	-.016	-.001
	Media	-.120	-.105	-.024	-.073	-.015	.019

Nota: n = tamaño muestral, RMSE = raíz media del error al cuadrado, Bias = Sesgo,  $\alpha$  = Coeficiente alfa,  $\omega$  = Coeficiente Omega, GLB = glb.algebraic, GLBa = glb.fa,  $\alpha$  or = Coeficiente alfa ordinal,  $\omega$  or = Coeficiente Omega ordinal; en negrita los valores superiores a  $|\cdot 05|$

#### 4 Categorías de respuesta

A continuación se presentan en la Tabla 7 y la Tabla 8 los resultados para el caso de ítems con cuatro opciones de respuesta. Para esta condición se observa una mejoría en la estimación de fiabilidad por parte de los coeficientes  $\alpha$  y  $\omega$  (RMSE medio de .083 y .068; con Sesgos medios de -.079 y -.063 respectivamente), sin embargo los valores siguen siendo elevados, manteniendo infraestimaciones en la recuperación del parámetro simulado. Por su

parte, el coeficiente GLB presenta valores más cercanos, aunque levemente superiores en algunos casos, a la fiabilidad simulada (RMSE medio de .036 y Sesgo medio de .01) produciéndose en algunos casos ligeras sobrestimaciones del parámetro simulado. En tanto, GLBa muestra ligeras infraestimaciones (RMSE medio de .050 y Sesgo medio de -.042), aunque no tan graves como las de  $\alpha$ . Al igual que en las dos condiciones previas, los coeficientes  $\alpha$  y  $\omega$  ordinal son los que muestran el mejor rendimiento, recuperando el parámetro de fiabilidad simulado de forma correcta para  $\omega$  ordinal y con un constante sesgo negativo para  $\alpha$  ordinal.

Tabla 7: 4 categorías 6 ítems congéntricos (fiabilidad = .731)

ítems asimétricos	n	as	cr	$\alpha$	$\omega$	GlB	glBa	$\alpha$ ordinal	$\omega$ ordinal
0	250	0.000	-0.385	.668	.683	.757	.723	.717	.733
	500	0.002	-0.292	.663	.677	.745	.706	.717	.732
	1000	0.001	-0.242	.657	.670	.735	.692	.717	.731
1	250	-0.191	-0.240	.664	.680	.755	.713	.717	.733
	500	-0.190	-0.159	.660	.675	.742	.692	.717	.732
	1000	-0.193	-0.114	.654	.668	.732	.674	.716	.731
2	250	-0.383	-0.093	.659	.676	.753	.706	.716	.733
	500	-0.383	-0.026	.655	.671	.740	.682	.717	.732
	1000	-0.386	0.013	.651	.665	.729	.663	.716	.731
3	250	-0.576	0.058	.654	.671	.749	.700	.716	.732
	500	-0.575	0.111	.651	.666	.737	.677	.716	.732
	1000	-0.579	0.141	.647	.661	.727	.657	.716	.730
4	250	-0.768	0.209	.649	.665	.745	.698	.715	.732
	500	-0.768	0.246	.647	.662	.732	.676	.716	.731
	1000	-0.772	0.268	.644	.658	.723	.659	.716	.730
5	250	-0.960	0.360	.646	.661	.743	.700	.715	.732
	500	-0.961	0.381	.645	.659	.732	.682	.715	.731
	1000	-0.965	0.396	.643	.657	.723	.668	.715	.730
6	250	-1.152	0.511	.648	.664	.745	.711	.714	.731
	500	-1.155	0.516	.648	.664	.737	.697	.715	.730
	1000	-1.159	0.524	.648	.663	.728	.686	.715	.730



Nota: n = tamaño muestral,  $\alpha$  = Coeficiente Alfa;  $\omega$  = Coeficiente Omega; GLB = Coeficiente GLB ("glb.fa"); GLBa = Coeficiente GLB ("glb.algebraic");  $\omega$  ordinal = Coeficiente Omega Ordinal;  $\alpha$  ordinal = Coeficiente Alfa ordinal

Tabla 8: RMSE y Sesgo de los coeficientes de fiabilidad para la condición de 4 categorías

ítems asimétricos	n	RMSE $\alpha$	RMSE $\omega$	RMSE GLB	RMSE GLBa	RMSE $\alpha$ or	RMSE $\omega$ or
0	250	<b>.070</b>	<b>.055</b>	.043	.029	.032	.026
	500	<b>.071</b>	<b>.057</b>	.036	.033	.025	.019
	1000	<b>.075</b>	<b>.062</b>	.032	.042	.020	.013
1	250	<b>.074</b>	<b>.058</b>	.041	.033	.033	.027
	500	<b>.074</b>	<b>.060</b>	.035	.045	.025	.020
	1000	<b>.078</b>	<b>.064</b>	.032	<b>.059</b>	.021	.013
2	250	<b>.078</b>	<b>.062</b>	.040	.039	.034	.027
	500	<b>.079</b>	<b>.063</b>	.035	<b>.054</b>	.026	.020
	1000	<b>.082</b>	<b>.067</b>	.033	<b>.070</b>	.021	.014
3	250	<b>.084</b>	<b>.067</b>	.040	.043	.035	.028
	500	<b>.083</b>	<b>.068</b>	.034	<b>.059</b>	.026	.020
	1000	<b>.086</b>	<b>.071</b>	.032	<b>.076</b>	.021	.014
4	250	<b>.089</b>	<b>.073</b>	.039	.046	.035	.028
	500	<b>.088</b>	<b>.072</b>	.034	<b>.060</b>	.027	.020
	1000	<b>.089</b>	<b>.075</b>	.033	<b>.075</b>	.022	.014
5	250	<b>.092</b>	<b>.077</b>	.040	.044	.036	.029
	500	<b>.090</b>	<b>.075</b>	.035	<b>.055</b>	.027	.020
	1000	<b>.089</b>	<b>.076</b>	.035	<b>.065</b>	.022	.014
6	250	<b>.091</b>	<b>.075</b>	.040	.039	.037	.030
	500	<b>.087</b>	<b>.071</b>	.036	.042	.028	.021
	1000	<b>.085</b>	<b>.070</b>	.032	.048	.022	.014
media		.083	.068	.036	.050	.027	.021
ítems asimétricos	n	Bias $\alpha$	Bias $\omega$	Bias GLB	Bias GLBa	Bias $\alpha$ or	Bias $\omega$ or
0	250	<b>-.063</b>	-.048	.026	-.008	-.014	.002
	500	<b>-.068</b>	<b>-.054</b>	.015	-.025	-.013	.002
	1000	<b>-.074</b>	<b>-.061</b>	.004	-.039	-.014	.000
1	250	<b>-.067</b>	<b>-.051</b>	.024	-.018	-.014	.002
	500	<b>-.071</b>	<b>-.056</b>	.011	-.039	-.014	.001
	1000	<b>-.077</b>	<b>-.063</b>	.001	<b>-.057</b>	-.014	.000
2	250	<b>-.072</b>	<b>-.055</b>	.022	-.025	-.014	.002
	500	<b>-.075</b>	<b>-.060</b>	.009	-.049	-.014	.001

	1000	<b>-.080</b>	<b>-.066</b>	-.002	<b>-.068</b>	-.014	.000
3	250	<b>-.077</b>	<b>-.060</b>	.018	-.031	-.015	.002
	500	<b>-.080</b>	<b>-.065</b>	.006	<b>-.054</b>	-.014	.001
	1000	<b>-.084</b>	<b>-.070</b>	-.004	<b>-.074</b>	-.015	.000
4	250	<b>-.082</b>	<b>-.066</b>	.014	-.033	-.015	.001
	500	<b>-.084</b>	<b>-.069</b>	.001	<b>-.055</b>	-.015	.000
	1000	<b>-.087</b>	<b>-.073</b>	-.008	<b>-.072</b>	-.015	-.001
5	250	<b>-.085</b>	<b>-.069</b>	.012	-.030	-.016	.001
	500	<b>-.086</b>	<b>-.072</b>	.001	-.049	-.016	.000
	1000	<b>-.088</b>	<b>-.074</b>	-.008	<b>-.063</b>	-.016	-.001
6	250	<b>-.083</b>	<b>-.066</b>	.014	-.020	-.017	.000
	500	<b>-.083</b>	<b>-.067</b>	.006	-.034	-.016	-.001
	1000	<b>-.083</b>	<b>-.068</b>	-.003	-.045	-.016	-.001
	media	-.079	-.063	.008	-.042	-.015	.001

Nota: n = tamaño muestral, RMSE = raíz media del error al cuadrado, Bias = Sesgo,  $\alpha$  = Coeficiente alfa,  $\omega$  = Coeficiente Omega, GLB = glb.algebraic, GLBa = glb.fa,  $\alpha$  or = Coeficiente alfa ordinal,  $\omega$  or = Coeficiente Omega ordinal; en negrita los valores superiores a |.05|

### 5 Categorías de respuesta

Finalmente en la Tabla 9 y Tabla 10 se muestran los resultados para la condición de cinco opciones de respuesta. Se observa una mejoría en la estimación de la fiabilidad por parte de los coeficientes  $\alpha$  y  $\omega$  frente a las condiciones anteriores. Sin embargo, en casi todas las condiciones se observan RMSE relativamente altos (valor medio de .068) y Sesgos negativos (con valor medio de -.064) para  $\alpha$ , mientras que para  $\omega$  los valores de RMSE fluctúan en torno a .053, logrando en condiciones con pocos ítems asimétricos valores más estables, mientras el Sesgo medio fue -.048, mostrando infraestimaciones ligeras en condiciones menos adversas. Con respecto a GLB y GLBa se observan valores medios de RMSE similares (en torno a .038) y sesgo medio positivo para el primero (valor de .019) y negativos para el segundo (-.027), por lo que el primero entrega ligeras sobrestimaciones del parámetro mientras el segundo infraestima ese valor. Al igual que en las condiciones anteriores, los coeficientes  $\alpha$  y  $\omega$  ordinal presentan una adecuada recuperación del parámetro

simulado, persistiendo el sesgo negativo en  $\alpha$  ordinal por la ausencia de tau-equivalencia en los datos.

Tabla 9: 5 categorías 6 ítems congéntricos (fiabilidad = .731)

ítems asimétricos	n	as	cr	$\alpha$	$\omega$	GLB	GLBa	$\alpha$ ordinal	$\omega$ ordinal
0	250	0.001	-0.322	.684	.699	.769	.737	.716	.732
	500	0.001	-0.225	.682	.696	.759	.723	.717	.732
	1000	0.001	-0.164	.677	.691	.749	.710	.716	.731
1	250	-0.160	-0.295	.680	.696	.766	.728	.708	.724
	500	-0.159	-0.214	.678	.693	.756	.710	.717	.732
	1000	-0.160	-0.161	.673	.688	.746	.694	.716	.731
2	250	-0.322	-0.267	.674	.691	.763	.722	.703	.718
	500	-0.319	-0.202	.673	.689	.753	.701	.716	.731
	1000	-0.322	-0.158	.669	.684	.744	.682	.716	.730
3	250	-0.482	-0.239	.668	.686	.758	.716	.700	.714
	500	-0.480	-0.189	.667	.683	.749	.694	.715	.730
	1000	-0.483	-0.156	.664	.679	.739	.676	.715	.730
4	250	-0.642	-0.211	.662	.679	.755	.712	.700	.714
	500	-0.641	-0.177	.662	.678	.746	.692	.714	.729
	1000	-0.644	-0.154	.659	.674	.736	.675	.715	.730
5	250	-0.802	-0.181	.657	.673	.751	.711	.705	.718
	500	-0.802	-0.167	.658	.673	.742	.695	.714	.729
	1000	-0.805	-0.152	.656	.671	.733	.681	.715	.730
6	250	-0.963	-0.154	.656	.673	.752	.717	.715	.729
	500	-0.962	-0.157	.658	.673	.742	.705	.715	.730
	1000	-0.966	-0.149	.658	.672	.734	.695	.715	.730

Nota: n = tamaño muestral,  $\alpha$  = Coeficiente Alfa;  $\omega$  = Coeficiente Omega; GLB = Coeficiente GLB ("glb.fa"); GLBa = Coeficiente GLB ("glb.algebraic");  $\omega$  ordinal = Coeficiente Omega Ordinal;  $\alpha$  ordinal = Coeficiente Alfa ordinal

Tabla 10: RMSE y Sesgo de los coeficientes de fiabilidad para la condición de 5 categorías

ítems asimétricos	n	RMSE $\alpha$	RMSE $\omega$	RMSE GLB	RMSE GLBa	RMSE $\alpha$ or	RMSE $\omega$ or
0	250	<b>.056</b>	.042	.051	.028	.032	.026
	500	<b>.053</b>	.040	.042	.022	.024	.019
	1000	<b>.056</b>	.043	.035	.026	.021	.013
1	250	<b>.059</b>	.044	.049	.028	.040	.032
	500	<b>.057</b>	.043	.040	.029	.026	.019
	1000	<b>.059</b>	.045	.033	.041	.021	.013
2	250	<b>.064</b>	.048	.047	.030	.045	.035
	500	<b>.062</b>	.047	.038	.037	.027	.020
	1000	<b>.064</b>	.049	.033	<b>.051</b>	.021	.013
3	250	<b>.070</b>	<b>.053</b>	.044	.033	.048	.038
	500	<b>.067</b>	<b>.051</b>	.037	.042	.027	.020
	1000	<b>.069</b>	<b>.053</b>	.032	<b>.058</b>	.021	.013
4	250	<b>.076</b>	<b>.059</b>	.043	.036	.046	.037
	500	<b>.073</b>	<b>.057</b>	.036	.045	.028	.021
	1000	<b>.073</b>	<b>.059</b>	.032	<b>.059</b>	.022	.014
5	250	<b>.081</b>	<b>.065</b>	.042	.037	.042	.034
	500	<b>.077</b>	<b>.062</b>	.035	.043	.028	.021
	1000	<b>.076</b>	<b>.062</b>	.032	<b>.053</b>	.053	.014
6	250	<b>.082</b>	<b>.066</b>	.042	.034	.034	.027
	500	<b>.077</b>	<b>.062</b>	.036	.035	.027	.020
	1000	<b>.075</b>	<b>.061</b>	.031	.039	.022	.014
media		.068	.053	.039	.038	.031	.022
ítems asimétricos	n	Bias $\alpha$	Bias $\omega$	Bias GLB	Bias GLBa	Bias $\alpha$ or	Bias $\omega$ or
0	250	-.047	-.032	.039	.006	-.015	.001
	500	-.049	-.035	.028	-.008	-.014	.001
	1000	<b>-.054</b>	-.040	.018	-.021	-.015	.000
1	250	<b>-.051</b>	-.035	.035	-.003	-.022	-.007
	500	<b>-.053</b>	-.038	.025	-.021	.014	.001
	1000	<b>-.058</b>	-.043	.016	-.037	-.015	.000
2	250	<b>-.056</b>	-.039	.032	-.009	-.028	-.013
	500	<b>-.058</b>	-.042	.022	-.030	-.015	.000
	1000	<b>-.062</b>	-.047	.013	-.049	-.015	-.001
3	250	<b>-.063</b>	-.045	.028	-.015	-.032	-.017
	500	<b>-.064</b>	-.047	.019	-.036	-.016	-.001
	1000	<b>-.067</b>	<b>-.052</b>	.008	<b>-.055</b>	-.015	-.001
4	250	<b>-.069</b>	<b>-.052</b>	.024	-.019	-.031	-.017
	500	<b>-.069</b>	<b>-.053</b>	.015	-.039	-.017	-.002
	1000	<b>-.072</b>	<b>-.057</b>	.005	<b>-.056</b>	-.016	-.001
5	250	<b>-.074</b>	<b>-.058</b>	.020	-.020	-.026	-.013

	500	<b>-.073</b>	<b>-.058</b>	.011	-.036	-.017	-.002
	1000	<b>-.074</b>	<b>-.060</b>	.002	<b>-.050</b>	-.016	-.001
6	250	<b>-.074</b>	<b>-.058</b>	.021	-.014	-.016	-.002
	500	<b>-.073</b>	<b>-.057</b>	.012	-.026	-.016	-.001
	1000	<b>-.073</b>	<b>-.058</b>	.003	-.036	-.016	-.001
	media	-.064	-.048	.019	-.027	-.017	-.004

Nota: n = tamaño muestral, RMSE = raíz media del error al cuadrado, Bias = Sesgo,  $\alpha$  = Coeficiente alfa,  $\omega$  = Coeficiente Omega, GLB = glb.algebraic, GLBa = glb.fa,  $\alpha$  or = Coeficiente alfa ordinal,  $\omega$  or = Coeficiente Omega ordinal; en negrita los valores superiores a |.05|

## Conclusión

A modo de conclusión se pueden mencionar lo siguiente: a) se observa un claro efecto del número de categorías de respuesta del ítem en la estimación de la fiabilidad sobre todo en los coeficientes  $\alpha$  y  $\omega$ , produciéndose infraestimaciones mayores (Sesgos negativos más altos) a medida que decrece el número de categorías; b) del mismo modo, se observa que la incorporación sucesiva de ítems asimétricos tiene un efecto negativo en la estimación de la fiabilidad produciéndose sesgos negativos para las versiones lineales de  $\alpha$  y  $\omega$  pero no así para sus versiones ordinales, las cuales en general recuperan de manera óptima el parámetro simulado; c) se observan las consecuencias del incumplimiento del supuesto de tau-equivalencia en la estimaciones entregadas tanto por el coeficiente  $\alpha$  como por su versión ordinal; d) los coeficientes GLB y GLBa mostraron un mejor desempeño (menor sesgo y RMSE) que  $\alpha$  y  $\omega$  lineales, llegando a mostrar valores aceptables en las condiciones de 3, 4 y 5 categorías; e) de todos los coeficientes evaluados el que mejor recupera el parámetro simulado es el coeficiente  $\omega$  ordinal.



## **ESTUDIO 2: ESTIMACIÓN DE LA FIABILIDAD EN MODELOS MULTIDIMENSIONALES**

Con el objetivo de determinar el sesgo de los coeficientes que asumen unidimensionalidad ( $\alpha$ ,  $\omega_t$  y GLB) en condiciones de multidimensionalidad (en presencia de correlaciones positivas entre errores y en modelos Bifactor), se mostrará cómo estos coeficientes tienden a generar sobre-estimaciones de la fiabilidad, entregando valores inflados que no corresponden al valor de la fiabilidad del factor general. En presencia de modelos unidimensionales con correlación entre los errores de los ítems es posible corregir la estimación de la fiabilidad mediante la fórmula de Raykov (2001) que, como se ha dicho anteriormente, corrige la correlación entre los errores incorporándolos en el denominador de la ecuación al estimar la fiabilidad. En casos donde existan múltiples correlaciones entre errores, la estructura puede ser analizada con modelos Bifactor en los cuales se plantea un factor general común a todos los ítems y factores grupales, específicos a algunos ítems (que representan, en este contexto, la causa de las correlaciones entre los errores de esos ítems), de este modo la propuesta de Raykov es extendida, realizando la estimación de la fiabilidad del factor general mediante el coeficiente  $\omega_h$ . Para ilustrar lo anteriormente descrito, este estudio estará dividido en tres apartados. En el primero, a modo de introducción, se muestra a través de sencillos ejemplos el efecto de la correlación positiva entre errores en la estimación de la fiabilidad, y como se corrige con la ecuación de Raykov. Por su parte, en el segundo apartado se presenta, mediante simulaciones bajo un modelo Bifactor, el sesgo que se produce al estimar la fiabilidad mediante los coeficientes que asumen unidimensionalidad, de esta forma se expande el caso sencillo de correlación entre los errores de algún par de ítems al caso más

complejo de un modelo Bifactor. Finalmente en el tercer apartado se presenta el caso de la redundancia entre ítems, que al ser evaluados mediante modelos Bifactor permite corregir los sesgos de sobrestimación de los coeficientes unidimensionales.

A continuación se presenta una serie de ejemplos con matrices muestrales de correlaciones simuladas, en el primer caso se analizará una matriz entre 5 ítems (con parámetros lambdas con valor de .50), posteriormente se añadirá un sexto ítem generado a partir del quinto (redundancia como fue definida por Cronbach y que supondría un problema de solapamiento de contenido), formando una correlación entre los errores de ambos ítems. Para finalizar estos ejemplos, se mostrará una extensión de la situación descrita considerando 9 ítems con varias correlaciones entre los errores de los ítems. Estos análisis se realizaron con el software R 3.2 y RStudio 0.98 mediante el paquete “*psych*” y la función “*omega*”

## EJEMPLOS DE ESTIMACIÓN DE LA FIABILIDAD CON CORRELACIÓN ENTRE ERRORES.

### **Caso 1. Modelo Unidimensional de 5 ítems**

Para este ejemplo la fiabilidad teórica simulada fue de .625 (5 ítems con lambdas fijados a un valor de .50), en la Tabla 11 se aprecia la matriz de correlación muestral. Sin embargo, la fiabilidad empírica, obtenida como la correlación al cuadrado entre la escala (sumatoria de los ítems) y el factor latente, es levemente menor .621. Del mismo modo, las estimaciones puntuales de  $\alpha$  y  $\omega_t$  obtienen un valor de .612 (siendo iguales al estar bajo un modelo tau-equivalente), levemente menores al parámetro simulado y a la fiabilidad empírica. Esto es explicado por fluctuaciones muestrales menores (la matriz población presenta correlaciones



de .25 entre todos los ítems). Como se observa, al no existir violación del supuesto de unidimensionalidad, los valores de  $\alpha$  y  $\omega_t$  son relativamente precisos, no existiendo sesgos graves.

Tabla 11: Matriz de correlación entre 5 ítems

	i1	i2	i3	i4	i5
i1	1	.2328562	.2226321	.2302269	.2345802
i2	.2328562	1	.2559890	.2420127	.2342596
i3	.2226321	.2559890	1	.2468956	.2549380
i4	.2302269	.2420127	.2468956	1	.2404937
i5	.2345802	.2342596	.2549380	.2404937	1

### **Caso 2: 6 ítems con correlación entre errores**

A la matriz anterior se incorporó un sexto ítem que presenta información redundante con el quinto, provocando una correlación entre errores debido a la presencia de covarianza más allá del factor general (ver Tabla 12). En esta situación, la fiabilidad teórica simulada (con el mismo valor de los parámetros lambda anteriores) fue de .666, en tanto la fiabilidad empírica (correlación al cuadrado entre la escala y el factor) fue de .590, valor que resulta menor al simulado debido a que existe mayor varianza error (atribuida al solapamiento entre el quinto y del sexto ítem y al hecho de que el sexto ítem no aporta información). Para esta condición el valor del coeficiente  $\alpha$  fue de .697 y el del coeficiente  $\omega_t$  fue de .72, ambos con claras sobreestimaciones del parámetro de fiabilidad simulado e incluso mayores si se considera la fiabilidad empírica. Al incorporar la corrección propuesta por Raykov se obtiene un valor de .588, virtualmente idéntico al de la fiabilidad empírica. Por lo tanto, en situaciones donde exista correlación entre errores se debe efectuar la corrección pertinente para evitar sesgos de sobrestimación del parámetro que para  $\alpha$  y  $\omega_t$  puede llegar a ser mayor a 1 décima cuando

solo existe correlación entre un par de errores. Si con este mismo conjunto de datos se estima el coeficiente  $\omega_h$  se obtiene un valor de .593, recuperando al igual que la corrección de Raykov adecuadamente el valor de la fiabilidad empírica.

Tabla 12: Matriz de correlación entre 6 ítems con correlación entre errores de ítem 5 y 6

	i1	i2	i3	i4	i5	i6
i1	1	.2328562	.2226321	.2302269	.2345802	.2082212
i2	.2328562	1	.2559890	.2420127	.2342596	.2174524
i3	.2226321	.2559890	1	.2468956	.2549380	.2219199
i4	.2302269	.2420127	.2468956	1	.2404937	.2134908
i5	.2345802	.2342596	.2549380	.2404937	1	.8991713
i6	.2082212	.2174524	.2219199	.2134908	.8991713	1

### Caso 3. 9 ítems y 3 factores específicos:

Para finalizar estos ejemplos, se presenta el caso de un test compuesto por 9 ítems (3 de los cuales presentan información y los otros 6 son obtenidos por redundancia): 1, 1a, 1b, 2, 2a, 2b y 3, 3a, 3b (la matriz de correlación se puede apreciar en Tabla 13). La fiabilidad simulada fue de .75 (con lambdas de .50 para los 9 ítems si es que los errores no estuviesen correlacionados). Sin embargo, la fiabilidad empírica (correlación al cuadrado entre la escala y el factor) presenta un valor mucho menor de .478. En esta situación si se estimara la fiabilidad mediante los coeficientes  $\alpha$  y  $\omega_t$  ajustando a un modelo unidimensional se produciría una gran sobrestimación (comparado tanto con la fiabilidad simulada como con la empírica), ambos coeficientes obtuvieron un valor de .842. Al controlar las múltiples correlaciones entre los errores de los ítems con la propuesta de Raykov (fórmula 16) la fiabilidad entrega un valor de .478, igual al de la fiabilidad empírica; sin embargo, a nivel

aplicado tiene una seria limitación, puesto que requiere identificar entre qué pares de errores se presentan correlaciones para posteriormente incorporarlas en la estimación de la fiabilidad. Para superar esa limitación el coeficiente  $\omega_h$  permite estimar, dentro de modelos multidimensionales, la fiabilidad del factor común controlando la varianza asociada a correlaciones entre errores o a factores grupales. En consecuencia, al contrastar con una estructura Bifactor utilizando la función “omega” del paquete “*psych*”, considerando un factor general común a todos los ítems y tres factores grupales específicos a algunos ítems que permiten explicar las correlaciones entre esos errores. Bajo esta estructura el valor del coeficiente  $\alpha$  se mantiene en .842, en tanto que  $\omega_t$  presenta un valor incluso mayor de .968 (esto se debe a que considera los lambdas del factor común y de los factores grupales, es decir, trabaja con la comunalidad) muy cercano a 1, lo que puede indicar, tal como es el caso, un alto solapamiento entre los ítems. En esta situación el coeficiente  $\omega_h$ , que estima la fiabilidad atribuida al factor general, entrega un valor de .470, muy similar al obtenido con el procedimiento de corrección por correlación entre errores y a la fiabilidad empírica. Por lo tanto, en situaciones de alta correlación entre los errores de varios ítems de un test, los coeficiente  $\alpha$  y  $\omega_t$  se vuelven estimadores positivamente sesgados, sobrestimando considerablemente la fiabilidad del factor general. En la Figura 5 se observa la representación de este modelo Bifactor, apreciándose una recuperación más adecuada de los parámetros lambda del factor general con valores en torno a .50.

Tabla 13: Matriz de correlación entre 9 ítems con múltiples correlaciones entre errores

	i1	i1a	i1b	i2	i2a	i2b	i3	i3a	i3b
i1	1	.8945348	.8953996	.2328562	.2026447	.2077678	.2226321	.2005137	.2054417
i1a	.8945348	1	.8006497	.2119352	.1823656	.1879057	.1996920	.1804817	.1845810
i1b	.8953996	.8006497	1	.2076573	.1835882	.1830204	.2072234	.1866136	.1891862
i2	.2328562	.2119352	.2076573	1	.8948944	.8974009	.2559890	.2314878	.2251568
i2a	.2026447	.1823656	.1835882	.8948944	1	.8028464	.2340458	.2083374	.2057940
i2b	.2077678	.1879057	.1830204	.8974009	.8028464	1	.2296221	.2079220	.2022303
i3	.2226321	.1996920	.2072234	.2559890	.2340458	.2296221	1	.9008274	.8990526
i3a	.2005137	.1804817	.1866136	.2314878	.2083374	.2079220	.9008274	1	.8107809
i3b	.2054417	.1845810	.1891862	.2251568	.2057940	.2022303	.8990526	.8107809	1

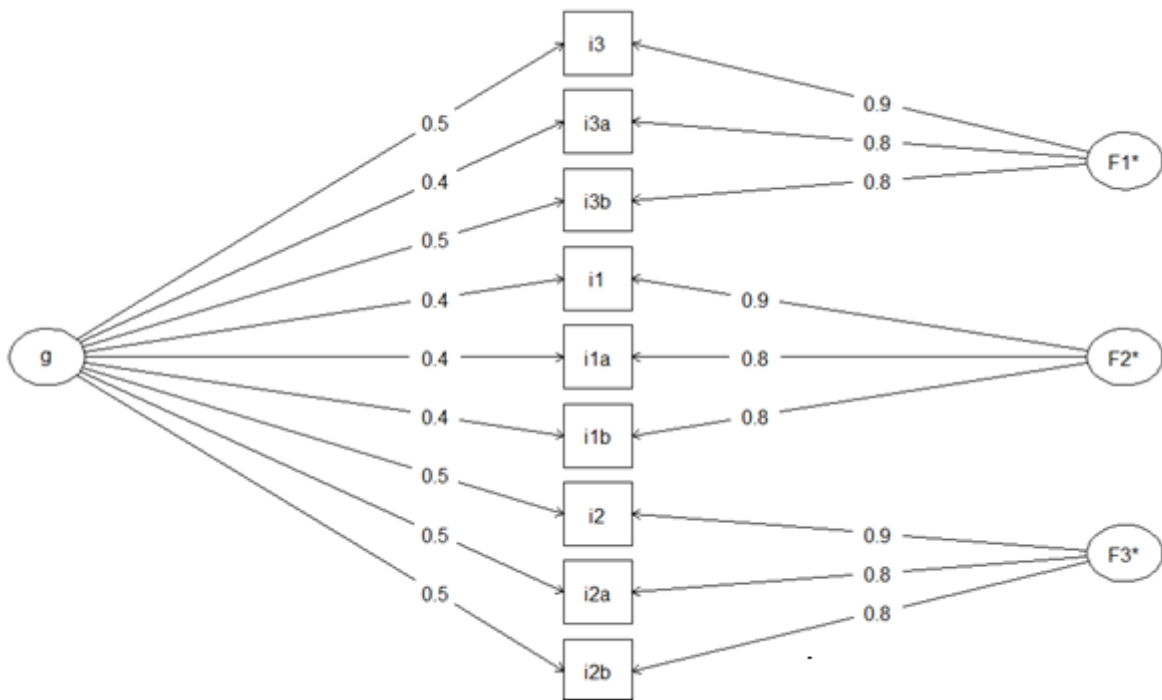


Figura 5: Modelo Bifactor para 9 ítems con 6 ítems redundantes.

## ESTUDIO 2.1: SESGOS EN LA ESTIMACIÓN DE LA FIABILIDAD EN MODELOS BIFACTOR

En este segundo estudio se amplía lo presentado anteriormente, generándose estructuras Bifactor en las cuales se evaluará la estimación de la fiabilidad que entregan los coeficientes  $\alpha$ ,  $\omega_t$ , GLB, GLBa y  $\omega_h$ .

### Método

#### *Generación de datos*

Los datos fueron generados usando los programas R 3.2 (R Development Core Team, 2013) y RStudio 0.98 (Racine, 2012), de acuerdo al siguiente modelo factorial:

$$X_{ij} = \sum_{k=1}^k \lambda_{jk} * F_k + \sqrt{(1 - \sum_{k=1}^k \lambda_{jk}^2)} * e_j \quad (34)$$

Donde  $X_{ij}$  es la respuesta simulada del sujeto  $i$  en el ítem  $j$ ,  $\lambda_{jk}$  es la saturación factorial del ítem  $j$  en el Factor  $k$  (el cual fue generado mediante un modelo Bifactor con un factor general y tres factores grupales),  $F_k$  corresponde al factor latente generado mediante una distribución normal estandarizada (media 0 y varianza 1), y  $e_j$  es el error de medida aleatorio de cada ítem siguiendo igualmente una distribución normal estandarizada.

#### *Condiciones simuladas*

Con el objetivo de evaluar el sesgo de los coeficientes de fiabilidad  $\alpha$ ,  $\omega_t$ , GLB (glb.fa), GLBa (glb.algebraic) y  $\omega_h$  en datos que incumplen el supuesto de Unidimensionalidad, se generó el modelo Bifactor con un Factor General y tres factores

grupales. Para evitar problemas de convergencia, se trabajó con un tamaño muestral de 1000 casos y una estructura con 18 ítems (6 ítems para cada factor grupal). En la Figura 6 se observa un ejemplo del modelo Bifactor simulado. Los parámetros lambda del factor general fueron variando desde .40 hasta .85 (.40, .45, .50, .55, .60, .65, .70, .75, .80, .85) en tanto los lambdas de los factores grupales se fijaron a un valor de .40. En cada condición la fiabilidad simulada del factor general fue aumentando a medida en que aumentaba el valor de los parámetros lambda (.637, .695, .744, .784, .818, .847, .871, .891, .909, .923, respectivamente). Se realizaron 500 réplicas por condición.

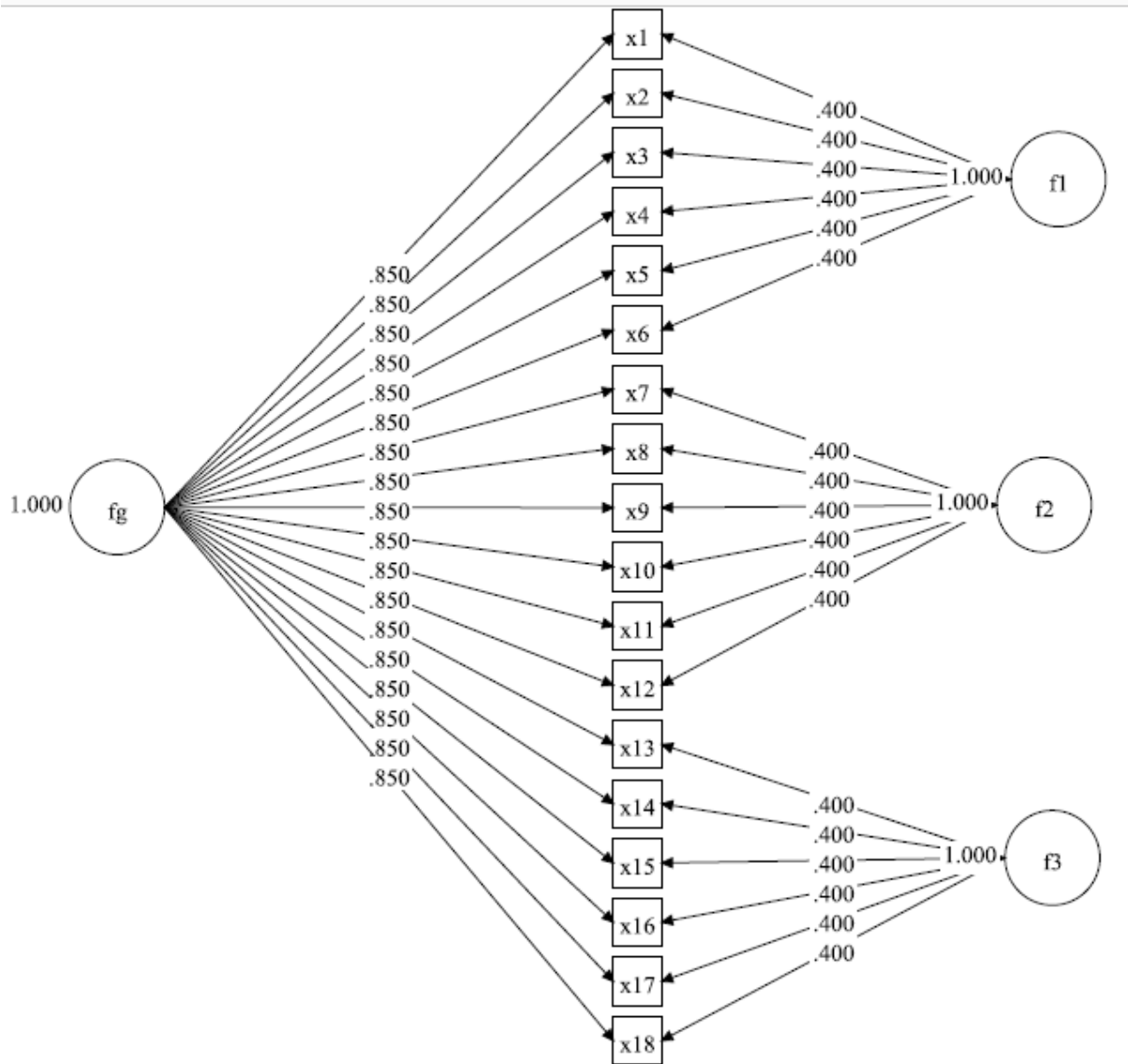


Figura 6: Ejemplo de modelo Bifactor poblacional simulado para la condición de fiabilidad = .923

#### *Análisis de datos*

Los principales análisis fueron realizados mediante los paquete *Psych* (Revelle, 2015b) y *GPArotation* (Bernaards & Jennrich, 2015), estos paquetes permiten estimar mediante el comando “omega” los coeficientes  $\alpha$ ,  $\omega_t$  y  $\omega_h$ , en tanto para la estimación de GLB se utilizó dos aproximaciones computacionales: *glb.fa* (Revelle, 2015a) y *glb.algebraic*

(Moltner & Revelle, 2015), este último algoritmo ya ha sido trabajado por autores como Hunt & Bentler (2015).

Con el fin de evaluar la precisión de los distintos estimadores en la recuperación de la fiabilidad, se calculó la raíz media del error al cuadrado (*RMSE*) y el sesgo (*bias*). Ambos han sido definidos previamente.

## Resultados

En la Tabla 14 se presentan las estimaciones ofrecidas por cada uno de los coeficientes evaluados para las 10 condiciones de fiabilidad simulada. Se aprecia que para valores bajos de fiabilidad del factor general, la sobrestimación que producen los coeficientes  $\alpha$ ,  $\omega_t$  y las dos versiones de GLB es muy elevada, por ejemplo si la fiabilidad atribuida al factor general es de .637, estos coeficientes entregan valores mayores a .80, lo que representan un gran sesgo positivo. En tanto el coeficiente  $\omega_h$  recupera de forma más precisa la fiabilidad con un valor de .618. En todas las condiciones, incluso cuando la fiabilidad del factor general era relativamente alta, los coeficientes  $\alpha$ ,  $\omega_t$ , GLB y GLBa presentaron siempre sobrestimaciones. Por lo que de todos ellos, el único que entrega estimaciones precisas de la fiabilidad es el coeficiente  $\omega_h$ . En la Figura 6 se observa la relación entre estos coeficientes y el valor de la fiabilidad en el eje x.



Tabla 14: Recuperación de la fiabilidad con diferentes estimadores

<b>Fiabilidad</b>	<b><math>\alpha</math></b>	<b><math>\omega_t</math></b>	<b>GLBfa</b>	<b>GLBa</b>	<b><math>\omega_h</math></b>
.637	.825	.856	.856	.874	.618
.695	.857	.879	.882	.898	.679
.744	.884	.903	.905	.918	.730
.784	.906	.923	.924	.935	.773
.818	.925	.939	.940	.949	.809
.847	.940	.953	.953	.960	.839
.871	.954	.966	.966	.971	.866
.891	.965	.975	.975	.979	.887
.909	.975	.984	.984	.987	.906
.923	.983	.991	.991	.993	.922

Nota:  $\alpha$  = Coeficiente Alfa;  $\omega_t$  = Coeficiente omega total; GLBfa = Greatest Lower Bound (glb.fa); GLBa = Greatest Lower Bound (glb.algebraic);  $\omega_h$  = Coeficiente omega Jerárquico

En relación al sesgo, en la Tabla 15 se presentan los valores de este estadístico para cada una de las condiciones simuladas. Se aprecia claramente que a medida que la fiabilidad simulada aumenta el sesgo de los coeficientes que asumen unidimensionalidad va disminuyendo (relación que se ve más clara en la Figura 7). Por ejemplo, en las condiciones de baja fiabilidad, los coeficientes que trabajan asumiendo unidimensionalidad generan sobrestimaciones con sesgos muy elevados. De todos ellos el que presenta el mayor sesgo es GLBa (.237, cuando la fiabilidad es de .637), mientras el coeficiente  $\alpha$  genera el menor sesgo (.187 en la misma condición). Respecto a la condición de más alta fiabilidad del factor general (.923) se observa que las estimaciones realizadas con los coeficientes que asumen unidimensionalidad presentan sesgos positivos menores (por ejemplo de GLBa obtiene un sesgo de .069, del mismo el coeficiente  $\alpha$  presenta un sesgo positivo más bajo .060) al estar ya al límite teórico del valor que puede alcanzar la fiabilidad (1). Por su parte, el coeficiente  $\omega_h$  presentó en general mejores resultados, recuperando relativamente bien la fiabilidad simulada en cada una de las condiciones.

Tabla 15: Sesgos en la estimación de la fiabilidad en modelos bifactor

<b>Fiabilidad</b>	<b>sesgo <math>\alpha</math></b>	<b>sesgo <math>\omega_t</math></b>	<b>sesgo GLB</b>	<b>sesgo GLBa</b>	<b>sesgo <math>\omega_h</math></b>
.637	.187	.213	.219	.237	-.019
.695	.178	.200	.204	.219	.013
.744	.140	.159	.161	.174	.014
.784	.122	.139	.140	.151	-.011
.818	.107	.121	.121	.131	-.009
.847	.094	.106	.106	.114	-.007
.871	.084	.095	.095	.100	-.005
.891	.074	.084	.084	.088	-.004
.909	.067	.076	.076	.078	-.002
.923	.060	.068	.068	.069	-.001

Nota:  $\alpha$  = Coeficiente Alfa;  $\omega_t$  = Coeficiente Omega Total; GLB = glb.fa; GLBa = glb.algebraic;  $\omega_h$  = Coeficiente Omega Jerárquico

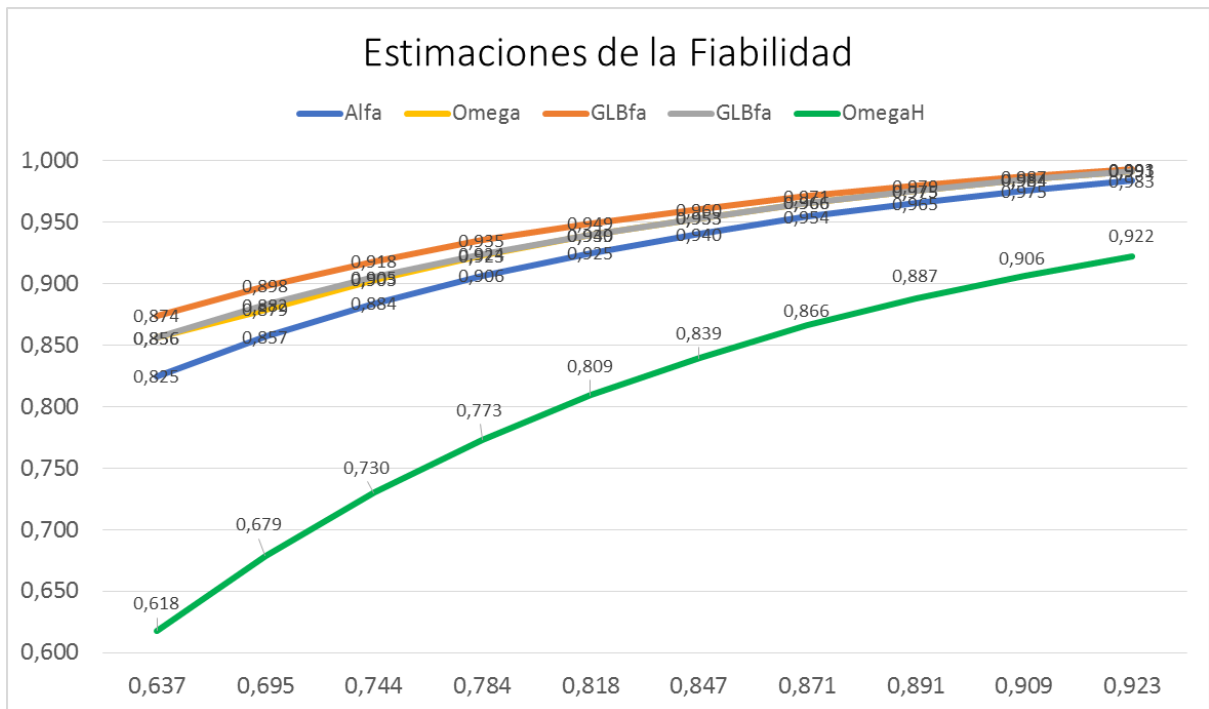


Figura 7: Estimaciones de la fiabilidad para el modelo Bifactor

### Conclusión

Bajo condiciones de multidimensionalidad con un factor general y tres factores grupales los coeficientes que estiman la fiabilidad del test asumiendo unidimensionalidad ( $\alpha$ ,  $\omega_t$ , GLB, GLBa) generan sobrestimaciones, incluso cuando los parámetros lambdas de los factores grupales tienden a ser relativamente bajos (.40) ya que no permiten identificar la varianza atribuida al factor general y controlar la debida a los factores grupales. Por lo tanto, en este tipo de condiciones, el único coeficiente que recupera adecuadamente la fiabilidad del factor general es  $\omega_h$ .

## ESTUDIO 2.2: SESGOS EN LA ESTIMACIÓN DE LA FIABILIDAD EN PRESENCIA DE ÍTEMS CON INFORMACIÓN REDUNDANTE

En este estudio se presenta el caso de ítems redundantes con errores positivamente correlacionados, estimándose la estructura Bifactor subyacente para así evaluar las estimaciones de la fiabilidad con los coeficientes  $\alpha$ ,  $\omega_t$ , GLB, GLBa y  $\omega_h$ .

### Método

#### *Generación de datos*

Los datos fueron generados usando los programas R 3.2 (R Development Core Team, 2013) y RStudio 0.98 (Racine, 2012), de acuerdo al siguiente modelo factorial:

$$X_{ij} = \sum_{k=1}^k \lambda_{jk} * F_k + \sqrt{(1 - \sum_{k=1}^k \lambda_{jk}^2)} * e_j \quad (37)$$

Donde  $X_{ij}$  es la respuesta simulada del sujeto  $i$  en el ítem  $j$ ,  $\lambda_{jk}$  es la saturación factorial del ítem  $j$  en el Factor  $k$  que corresponde al factor latente generado mediante una distribución normal estandarizada (media 0 y varianza 1), y  $e_j$  es el error de medida aleatorio de cada ítem siguiendo igualmente una distribución normal estandarizada.

#### *Condiciones simuladas*

Con el objetivo de evaluar el sesgo de los coeficientes de fiabilidad  $\alpha$ ,  $\omega_t$ , GLB, GLBa y  $\omega_h$  en datos que incumplen el supuesto de Unidimensionalidad, se generó el modelo unidimensional con ítems redundantes. De este modo se planteó una estructura de 3 ítems con información (con saturaciones directa en el factor general) que generaban 12 ítems

redundantes (4 ítems redundantes que se construían a partir de cada uno de los ítems informativos). Para evitar problemas de convergencia, se trabajó con un tamaño muestral de 1000. Los parámetros lambda del factor general fueron variando desde .30 hasta .85 (.30, .35, .40, .45, .50, .55, .60, .65, .70, .75, .80, .85), por su parte los valores de los lambdas de los ítems redundantes fueron de .70, .80 y .90, indicando alto solapamiento en el contenido de la información de esos ítems. En cada condición la fiabilidad simulada del factor general fue aumentando en la medida en que se aumentaba el valor de los parámetros lambda. Se realizaron 500 réplicas por condición.

### *Análisis de datos*

Los principales análisis fueron realizados mediante los paquete *Psych* (Revelle, 2015b) y *GPArotation* (Bernaards & Jennrich, 2015), estos paquetes permiten estimar mediante el comando “omega” los coeficientes  $\alpha$ ,  $\omega_t$  y  $\omega_h$ , en tanto para la estimación de GLB se utilizó dos aproximaciones computacionales: *glb.fa* (Revelle, 2015a) y *glb.algebraic* (Moltner & Revelle, 2015).

Con el fin de evaluar la precisión de los distintos estimadores en la recuperación de la fiabilidad, se calculó la raíz media del error al cuadrado (*RMSE*) y el sesgo (bias).

## **Resultados**

En la Tabla 16 se observan las estimaciones ofrecidas por los coeficientes de fiabilidad para las condiciones simuladas. En todos los casos, los coeficientes que asumen unidimensionalidad generan sobrestimaciones que para las condiciones más extremas (bajos

lambdas en el factor general y fiabilidades menores a .40), ofrecen valores superiores a .70 (para  $\alpha$ ) e incluso superiores a .80 ( $\omega_t$ , GLB, GLBa). De hecho, cuando aumenta la correlación entre los errores de los ítems (.80 y .90) estos coeficientes presentan valores superiores a .90, aun cuando la fiabilidad del factor general era relativamente baja, lo que indica efectivamente el efecto de la redundancia de la información en estos coeficientes. En contraste, se observa que la estimación de  $\omega_h$  recupera adecuadamente el parametro de la fiabilidad del factor general. Se observa que incluso en situaciones donde los lambdas del factor general son altos, los coeficientes que asumen unidimensionalidad siguen presentando valores elevados por encima de 1 décima.

Tabla 16: Estimaciones de la fiabilidad del factor general obtenidas para las condiciones simuladas

ítem red	Lambda		$\alpha$	$\omega_t$	GLBfa	GLBa	$\omega_h$
	FG	Fiabilidad					
.70	.30	.189	.703	.819	.821	.839	.215
	.35	.246	.721	.828	.828	.846	.254
	.40	.306	.739	.836	.836	.853	.307
	.45	.366	.757	.844	.844	.861	.365
	.50	.428	.775	.853	.853	.868	.425
	.55	.488	.793	.861	.861	.876	.483
	.60	.546	.810	.869	.869	.883	.541
	.65	.603	.827	.877	.877	.890	.597
	.70	.657	.843	.884	.884	.897	.649
	.75	.708	.858	.891	.891	.903	.700
	.80	.756	.872	.898	.898	.909	.746
	.85	.801	.885	.904	.905	.914	.790
.80	.30	.206	.768	.893	.894	.905	.228
	.35	.267	.783	.899	.899	.909	.274
	.40	.330	.799	.904	.904	.914	.333
	.45	.395	.815	.909	.909	.919	.395
	.50	.459	.831	.914	.914	.924	.458
	.55	.521	.847	.920	.920	.928	.520
	.60	.582	.862	.925	.925	.933	.579
	.65	.639	.876	.929	.929	.937	.636
	.70	.694	.890	.934	.934	.941	.691
	.75	.745	.903	.938	.938	.945	.742
	.80	.794	.915	.942	.942	.948	.789
	.85	.839	.926	.946	.946	.952	.834
.90	.30	.220	.819	.953	.953	.958	.237
	.35	.284	.833	.955	.955	.960	.291
	.40	.350	.847	.958	.958	.962	.353
	.45	.417	.861	.960	.960	.964	.418
	.50	.483	.875	.962	.962	.966	.483
	.55	.547	.889	.965	.965	.969	.547
	.60	.608	.902	.967	.967	.971	.608
	.65	.667	.914	.969	.969	.973	.666
	.70	.722	.926	.971	.971	.974	.721
	.75	.773	.937	.973	.973	.976	.772
	.80	.821	.947	.975	.975	.978	.820
	.85	.866	.956	.977	.977	.979	.865

Nota: ítem red = correlación simulada de los ítems redundantes; lambda FG = Saturación simulada en el Factor General,  $\alpha$ = Coeficiente Alfa;  $\omega_t$  = Coeficiente Omega Total; GLB = glb.fa; GLBa = glb.algebraica;  $\omega_h$  = Coeficiente Omega Jerárquico

En la Tabla 17 se presentan el Sesgo (RMSE obtiene el mismo valor excepto para  $\omega_h$ ) de los estimadores de la fiabilidad. Se observan sesgos positivos muy altos superiores a .50 para los cuatro coeficientes que asumen unidimensionalidad, en los casos donde la fiabilidad del factor general es muy pequeña. Esos sesgos van decreciendo a medida que aumenta la fiabilidad del factor general sin llegar a desaparecer. En cada condición el coeficiente  $\alpha$  es el que presenta el menor sesgo positivo de todos y GLBa el mayor. En contraste el coeficiente  $\omega_h$  presenta sesgos cercanos a 0, indicando una adecuada recuperación del parámetro.



Tabla 17: Sesgo de los coeficientes de fiabilidad en condiciones de ítem redundantes

ítem red	Lambda FG	Fiab	bias $\alpha$	bias $\omega_t$	bias GLB	bias GLBa	bias $\omega_h$	RMSE $\omega_h$
.7	.30	.189	.514	.630	.632	.649	.025	.064
	.35	.246	.475	.582	.582	.600	.008	.045
	.40	.306	.433	.530	.530	.548	.002	.04
	.45	.366	.390	.478	.478	.494	-.001	.037
	.50	.428	.347	.425	.425	.441	-.003	.033
	.55	.488	.305	.373	.373	.388	-.004	.03
	.60	.546	.264	.323	.323	.336	-.005	.027
	.65	.603	.224	.274	.274	.287	-.006	.024
	.70	.657	.186	.227	.227	.240	-.007	.021
	.75	.708	.150	.183	.183	.195	-.008	.019
	.80	.756	.116	.142	.142	.153	-.01	.017
	.85	.801	.084	.103	.103	.113	-.012	.017
.8	.30	.206	.561	.687	.688	.699	.022	.061
	.35	.267	.516	.632	.632	.642	.007	.044
	.40	.330	.469	.573	.573	.584	.003	.039
	.45	.395	.421	.514	.514	.524	.001	.035
	.50	.459	.373	.456	.456	.465	-.001	.031
	.55	.521	.326	.398	.398	.407	-.002	.028
	.60	.582	.281	.343	.343	.351	-.002	.024
	.65	.639	.237	.290	.290	.298	-.003	.021
	.70	.694	.196	.240	.240	.247	-.003	.018
	.75	.745	.158	.193	.193	.199	-.004	.015
	.80	.794	.121	.149	.149	.155	-.004	.013
	.85	.839	.088	.107	.107	.113	-.005	.011
.9	.30	.220	.599	.733	.734	.738	.018	.055
	.35	.284	.549	.671	.671	.676	.007	.043
	.40	.350	.497	.608	.608	.612	.003	.038
	.45	.417	.444	.543	.543	.548	.002	.034
	.50	.483	.392	.480	.480	.484	.001	.03
	.55	.547	.342	.418	.418	.422	0	.026
	.60	.608	.293	.359	.359	.362	0	.022
	.65	.667	.247	.303	.303	.306	-.001	.019
	.70	.722	.204	.250	.250	.253	-.001	.016
	.75	.773	.163	.200	.200	.203	-.001	.013
	.80	.821	.126	.154	.154	.156	-.001	.01
	.85	.866	.091	.111	.111	.113	-.001	.008

Nota: **ítem red** = correlación simulada de los ítems redundantes; Lambda FG = Saturación simulada en el Factor General, bias  $\alpha$  = sesgo Alfa; bias  $\omega_t$  = sesgo Omega Total; bias GLB = sesgo glb.fa; bias GLBa = sesgo glb.algebraic; bias  $\omega_h$  = sesgo Omega Jerárquico; RMSE  $\omega_h$  = RMSE omega Jerárquico

## **Conclusión**

Cuando existe redundancia entre un conjunto de ítems, los coeficientes que asumen unidimensionalidad generan grandes sobrestimaciones de la fiabilidad, tanto es así que incluso en situaciones donde la fiabilidad del factor general es débil ( $< .50$ ) estos coeficientes entregan valores sesgados elevados cercanos a .90, indicando una alta fiabilidad en donde realmente no la hay. En consecuencia el coeficiente  $\omega_h$  resulta apropiado para recuperar la fiabilidad atribuida al factor general controlando la varianza debida a la correlación entre errores.





### **ESTUDIO 3: ESTIMACIONES DE LA FIABILIDAD CON DATOS EMPÍRICOS**

El tercer estudio de esta tesis mostrará 2 ejemplos empíricos (con dos instrumentos diferentes) en los cuales se analizará en primer lugar la estructura factorial del cuestionario por medio de Análisis Factoriales Confirmatorios para posteriormente evaluar las evidencias de fiabilidad que se obtienen con los coeficientes de fiabilidad presentados en el apartado teórico.

#### **ESTUDIO 3.1: RE-ANÁLISIS DE LAS EVIDENCIAS DE FIABILIDAD DEL CECSCE.**

Para el primer caso se analizarán los datos del estudio de Gálvez-Nieto, Vera-Bachman y Trizano (2015). En ese artículo se presentaron evidencias de validez y fiabilidad del Cuestionario para evaluar clima social del centro escolar (CECSCE) en una muestra chilena ( $n = 1394$ ), sin embargo la evidencia de fiabilidad presentadas en ese artículo solo se limita a mostrar el valor obtenido con el coeficiente  $\alpha$ , por lo tanto resulta necesario, a la luz de los antecedentes teóricos y empíricos exhibidos en los apartados anteriores, realizar nuevas estimaciones de la fiabilidad, teniendo en cuenta la naturaleza de los datos, sus características distribucionales, así como la estructura subyacente a los mismos.

### **Método**

#### *Instrumento*

El CECSCE fue desarrollado originalmente en España por el equipo de Trianes (Trianes, Blanca, Morena, & Raya, 2006). Es un instrumento de autoinforme que evalúa el clima social escolar desde la perspectiva de los estudiantes. La escala está compuesta por 14

ítems con cinco categorías de respuesta (1 = muy en desacuerdo, 5 = muy de acuerdo). En el estudio original mostró evidencias de una estructura bidimensional correlacionada (Clima referente al Centro y Clima referente al Profesorado), sin embargo este análisis solo fue realizado con el método de Componentes Principales con rotación oblimin. Respecto a las evidencias de fiabilidad (por consistencia interna), el análisis fue realizado con el coeficiente  $\alpha$  de Cronbach (Clima Centro = .77 y Clima referente al Profesorado = .72). En Chile se presentaron evidencias de validez por Guerra et al. (2011), reportando una solución unidimensional (Clima del centro escolar) que explicó el 35.65% de la varianza total, y una segunda, bidimensional (Clima social referente al centro escolar y referente al profesorado), que explicó 45.29% de la varianza, estos análisis también se realizaron con Componentes Principales y oblimín. Respecto a la fiabilidad, los análisis se realizaron con el coeficiente  $\alpha$  de Cronbach (general = .86; Clima referente al Centro, = .83; factor Profesores = .72). Finalmente en el estudio de Gálvez-Nieto et al. (2015), las evidencias de validez fueron realizadas con el método de validación cruzada, con la primera muestra se realizaron Análisis Factoriales con ejes principales y rotación oblimin y con la segunda muestra se ratificó la estructura mediante un Análisis Factorial Confirmatorio, utilizando el método de máxima verosimilitud robusta, obteniéndose una estructura bidimensional con adecuados índices de bondad de ajuste ( $SB-\chi^2(76) = 290.77, p < .001$ ; TLI = 0.94; CFI = .95; RMSEA = .05, IC 90% [.04, .05].). Las evidencias de fiabilidad de consistencia interna nuevamente se limitaron al coeficiente  $\alpha$  (Clima Centro = .80 y Clima Profesores = .73)

## *Análisis de datos*

En primer lugar se presentarán los estadísticos descriptivos de centro, dispersión y forma de las 14 variables. Posteriormente se realizarán evaluaciones de las evidencias de validez contrastando los siguientes modelos: M1a Unidimensional (esencialmente) tau-equivalente, fijando las saturaciones factoriales; M1b Unidimensional congénico; M2 Bidimensional correlacionado; M3 Bifactor. Para cada modelo se trabajará con la matriz producto-momento de Pearson, estimando los parámetros con el método de Máxima Verosimilitud Robusta MLM, adecuado para datos no normales (Muthén & Muthén, 1998-2010); además se incorporará métodos para el análisis de datos categóricos mediante el uso de la matriz policórica (Asun et al., 2016; Flora & Curran, 2004; Jöreskog, 1994) utilizando el método de estimación “*Weighted Least Squares Mean and Variance Adjuste*” (mínimos cuadrados ponderados con media y varianza ajustada; WLSMV), adecuado para datos ordinales (Muthén & Muthén, 1998-2010). Se evaluó la calidad del modelo de reproducir los datos utilizando los siguientes índices de bondad de ajuste; chi-cuadrado de Satorra Bentler ( $SB-\chi^2$ ; Satorra & Bentler, 2001), índice de ajuste comparativo (CFI), índice de Tucker-Lewis (TLI) y error cuadrático medio de aproximación (RMSEA). Para los primeros índices, CFI y TLI, se considera un ajuste adecuado del modelo los valores superiores a .90 (Schumacher & Lomax, 1996) o, con un criterio más exigente, valor superiores a .95 (Schreiber, Nora, Stage, Barlow, & King, 2006), mientras que para el RMSEA se considera un ajuste razonable los valores inferiores a .08 (Browne & Cudeck, 1993) o .06 (Schreiber et al., 2006). Posteriormente se comparó los modelos entre sí, para los métodos basados en matrices de Pearson MLM se estimó la diferencia de  $SB-\chi^2$  utilizando el método de factor de corrección (Asparouhov & Muthén, 2013), por su parte para el modelo ordinal, mediante

matrices policóricas, se utilizó la opción DIFFTEST. Estos análisis fueron realizados con Mplus 7.1. Finalmente se evaluaron las evidencias de fiabilidad utilizando los siguientes estimadores:  $\alpha$  de Cronbach,  $\omega$ , GLB y GLBa,  $\alpha$  ordinal y  $\omega$  ordinal. El tamaño muestral fue de 1394 casos.

## Resultados

En la Tabla 18 se muestran los estadísticos descriptivos de centro, dispersión y forma de los ítems, considerando la media, D.E., asimetría y curtosis. Se observa que los ítems tienden a presentar asimetrías negativas, de hecho al estandarizarlos (dividiendo la asimetría por su error típico) solo los ítems 6 y 11 resultaron simétricos al presentar valores cercanos a 0 (todos los demás fueron superiores a |3|). Del mismo modo todos los ítems presentan curtosis negativas, con el valor estandarizado más cercano a 0 para el ítem 8. En esta línea, al evaluar la normalidad multivariante el estadístico de Mardia de Kurtosis multivariada resultó significativo ( $p < .001$ ). Reafirmandose la decisión de haber utilizado métodos de estimación robustos (MLM).

Tabla 18: Estadísticos descriptivos de los ítems del CECSC

Ítems	Media	Desv. típ.	Asimetría	Error típico	As Estand	Curtosis	Error típico	Cr Estand
Ítem1	3.77	1.04	-0.43	0.07	-6.55	-0.63	0.13	-4.79
Ítem2	3.69	1.07	-0.46	0.07	-6.95	-0.53	0.13	-4.04
Ítem3	3.77	0.96	-0.47	0.07	-7.05	-0.27	0.13	-2.08
Ítem4	3.54	1.19	-0.38	0.07	-5.68	-0.81	0.13	-6.15
Ítem5	3.74	1.17	-0.66	0.07	-9.98	-0.48	0.13	-3.69
Ítem6	3.08	1.21	0.01	0.07	0.14	-0.92	0.13	-7.02
Ítem7	3.26	0.98	-0.10	0.07	-1.47	-0.32	0.13	-2.47
Ítem8	3.92	1.13	-0.83	0.07	-12.61	-0.19	0.13	-1.47
Ítem9	3.89	1.17	-0.81	0.07	-12.29	-0.31	0.13	-2.40



Ítem10	3.67	1.18	-0.61	0.07	-9.17	-0.52	0.13	-3.96
Ítem11	2.99	1.15	0.03	0.07	0.47	-0.74	0.13	-5.63
Ítem12	3.63	1.18	-0.53	0.07	-8.08	-0.62	0.13	-4.71
Ítem13	3.73	1.19	-0.63	0.07	-9.47	-0.55	0.13	-4.18
Ítem14	2.58	1.40	0.35	0.07	5.36	-1.17	0.13	-8.90

Nota: Desv. típ = Desviación Típica ; As Estand = Coeficiente de asimetría estandarizado; Cr Estand = Coeficiente de curtosis estandarizado

En la Tabla 19 se presentan los índices de ajuste de los modelos contrastados. Se aprecia claramente que el modelo más restrictivo M1a, fijando los parámetros lambda para que tengan el mismo valor, genera los peores valores. Por su parte el modelo congénico M1b presenta índices de ajuste aceptables en ambas condiciones (Pearson y policóricas). El modelo bidimensional M2 igualmente presentó índices de ajuste satisfactorios con ambas matrices. Finalmente el modelo Bifactor M3 presentó los mejores índices de ajuste de todos los modelos evaluados tanto con matrices Pearson como Policóricas. Se observa que los valores de los índices comparativo CFI y TLI tienden a presentar mejor ajuste con datos categóricos, mientras RMSEA muestra un comportamiento inverso (menor ajuste en datos categóricos).

Tabla 19: Ajuste de los modelos CECSC

M	Correlaciones de Pearson (MLM)						Correlaciones policóricas (WLSMV)					
	SB- $\chi^2$	gl	P	CFI	TLI	RMSEA	$\chi^2$	gl	p	CFI	TLI	RMSEA
M1a	771,603	90	< .001	.840	.838	.074 (.069-.079)	1.395,854	90	< .001	.879	.877	.102 (.097-.107)
M1b	416	77	<.001	.920	.906	.056 (.051-.062)	623,092	77	< .001	.949	.940	.071 (.066-.077)
M2	294	76	< .001	.949	.939	.045 (.040-.051)	427	76	< .001	.967	.961	.058 (.052-.063)
M3	193	63	< .001	.969	.956	.038 (.032-.045)	252	63	< .001	.982	.975	.046 (.040-.052)

Nota: MLM = Máxima Verosimilitud Robusta; WLSMV = Mínimos Cuadrados Ponderados con media y varianza ajustada; M = Modelos; M1a = Modelos Tau-equivalente; M1b = Modelo Congénico; M2: Modelo bidimensional correlacionado; M3 = Modelo Bifactor. SB- $\chi^2$ = Satorra-Bentler Chi Cuadrado; gl = grados de libertad; CFI = índice de ajuste comparativo; TLI = Índice de Tucker-Lewis; RMSEA = Raíz del error cuadrático medio de aproximación;  $\chi^2$  = Chi-cuadrado

Se presenta en la Tabla 20 la comparación de los modelos basados en correlaciones de Pearson con método de estimación MLM mediante el factor de corrección y en la opción DIFFTEST para el método WLSMV estimado a partir de la matriz policórica. Se aprecia que de todos ellos, el que mejor ajuste presenta es el M3 (Bifactor, lamentablemente por un problema de estimación se presenta un caso Heywood en el ítem 13, por lo que este modelo no será contemplado para estimar la fiabilidad). Considerando lo expuesto anteriormente, a continuación se presentará los parámetros lambda obtenidos para posteriormente estimar la fiabilidad de este instrumento utilizando el modelo M2.

Tabla 20: Diferencia entre los modelos mediante el factor de corrección para MLM o mediante DIFFTEST para WLSMV

Pearson MLM	Modelo	$\Delta SB-\chi^2$	$\Delta df$	p
	M1a - M1b	410.380	13	< .001
	M1b - M2	116	1	< .001
	M2 - M3	100	13	< .001
Policóricas WLSMV	Modelo	$\Delta \chi^2$	$\Delta df$	p
	M1a-M1b	468.029	13	< .001
	M1b-M2	112	1	< .001
	M2-M3	159	13	< .001

Nota: MLM = Máxima Verosimilitud Robusta; WLSMV = Mínimos Cuadrados Ponderados con media y varianza ajustada; M1a = Modelos Tau-equivalente; M1b = Modelo Congénérico; M2: Modelo bidimensional correlacionado; M3 = Modelo Bifactor.

En la Tabla 21 se aprecian las saturaciones factoriales (parámetros de discriminación lambda junto a su error estándar) para los modelos M1b y M2 obtenidos por los métodos MLM y WLSMV con matrices de correlación producto-momento de Pearson y policóricas respectivamente. Para el caso unidimensional congénérico se observan buenos valores de lambda en todos los ítems, excepto en el ítem 3 donde se presenta el valor más bajo en ambas

matrices). Para el caso bidimensional se aprecian valores de lambda relativamente más altos en sus respectivos factores: Clima Social Escolar (ítems 1, 5, 6, 7, 9, 10, 11, 12) y Clima Social Profesores (ítems 2, 3, 4, 8, 13, 14), excepto para el ítem 3, que nuevamente presenta el valor más bajo.

Tabla 21: Saturaciones factoriales de los modelos contrastados

Matriz (Método)		Pearson (MLM)			Policóricas (WLSMV)	
Modelo	M1b	M2		M1b	M2	
Factor:	CS	CSE	CSP	CSE	CSE	CSP
Ítem	lambda (S.E)	lambda (S.E)	lambda (S.E)	lambda (S.E)	lambda (S.E)	lambda (S.E)
X1	.507 (.024)	.488 (.025)		.544 (.021)	.554 (.022)	
X2	.657 (.018)		.695 (.019)	.697 (.016)		.746 (.017)
X3	.299 (.028)		.325 (.029)	.323 (.025)		.342 (.026)
X4	.552 (.024)		.580 (.025)	.587 (.020)		.623 (.021)
X5	.525 (.023)	.547 (.023)		.560 (.020)	.575 (.020)	
X6	.671 (.017)	.684 (.017)		.701 (.015)	.718 (.015)	
X7	.457 (.026)	.466 (.026)		.485 (.021)	.498 (.022)	
X8	.536 (.022)		.582 (.023)	.594 (.020)		.632 (.021)
X9	.447 (.025)	.469 (.024)		.485 (.022)	.499 (.023)	
X10	.521 (.023)	.532 (.023)		.548 (.020)	.562 (.020)	
X11	.694 (.018)	.721 (.017)		.731 (.015)	.750 (.015)	
X12	.699 (.017)	.723 (.016)		.746 (.014)	.765 (.014)	
X13	.547 (.023)		.597 (.023)	.591 (.020)		.629 (.020)
X14	.536 (.020)		.579 (.020)	.589 (.019)		.627 (.020)

Nota: MLM = Máxima Verosimilitud Robusta; WLSMV = Mínimos Cuadrados Ponderados con media y varianza ajustada; M1b = Modelo Congénico; M2: Modelo bidimensional correlacionado;

### *Evidencias de fiabilidad*

Una vez evaluado el modelo factorial que mejor se ajusta a los datos se procede a obtener evidencias de fiabilidad. En la Tabla 22 se presentan las estimaciones obtenidas para M2. Se aprecia que el coeficiente  $\alpha$  siempre presenta un valor menor al obtenido con los

otros coeficientes, mostrando infraestimación al no cumplir el supuesto de continuidad (ya que las variables son categóricas) y el de tau-equivalencia, ya que como se observó previamente, los parámetros de discriminación obtenidos no son iguales para todos los ítems. En ambos factores, los coeficientes GLB presentan el valor más alto, seguido de cerca por  $\omega$  ordinal. Incluso considerando los intervalos de confianza de los coeficientes  $\alpha$  y  $\omega$ , en sus límites superiores nunca se encuentran los valores de los coeficientes GLB ni el de los coeficientes ordinales.

Tabla 22: Evidencias de fiabilidad para CECSCE basado en M2

Estimador\Factor	Clima Profesores	Clima Centro
$\alpha$ clásico	.733	.800
$\omega$ clásico	.742	.806
GLB	.787	.836
GLBa	.771	.835
$\alpha$ ordinal	.769	.825
$\omega$ ordinal	.776	.829

Nota:  $\alpha$  = coeficiente alfa;  $\omega$  = coeficiente omega; GLB = Greatest Lower Bound (glb.fa); GLBa = Greatest Lower Bound (glba.algebraic);  $\alpha$  ordinal = coeficiente alfa ordinal;  $\omega$  ordinal = coeficiente omega ordinal

## Conclusión

Considerando como correcta la estructura bidimensional, las estimaciones de la fiabilidad de cada factor realizadas con el coeficiente  $\alpha$  generan infraestimaciones al trabajar sobre datos congéntricos (violando el supuesto de tau-equivalencia), con asimetría negativa (incumpliendo el de normalidad) y categóricos (incumpliendo el de continuidad y linealidad), por lo tanto obtiene el menor valor posible de todos los coeficientes. En consecuencia se

deben utilizar coeficientes más apropiados, en este caso los valores de  $\omega$  ordinal y las alternativas de GLB tienden a coincidir, siendo coeficientes más robustos que  $\alpha$  al incumplimiento de los supuestos mencionados, resultando por lo tanto mejores estimaciones de la fiabilidad de cada factor.

### ESTUDIO 3.2: RE-ANÁLISIS DE LAS EVIDENCIAS DE FIABILIDAD DEL AAI-A

Para este ejemplo se analizarán los datos presentados en el estudio de Gálvez-Nieto, Vera-Bachman, Trizano, y García (2015). En ese trabajo los autores mostraron evidencias de validez y fiabilidad con el AAI-A.

## **Método**

### *Instrumento*

La Escala de Actitud hacia la Autoridad Institucional en Adolescentes (AAI-A) fue construida por el equipo de Cava (Cava, Estévez, Buelga, & Musitu, 2013) es una escala de autorreporte que evalúa las actitudes de los adolescentes hacia la autoridad en el contexto escolar. En el estudio de Cava et al (2013) se realizaron análisis factoriales exploratorios, confirmatorios (con métodos robustos) y multigrupo en España y México, mostrando adecuados índices de ajuste. La escala consta de 9 ítems tipo Likert con cuatro categorías de respuesta (1 = nada de acuerdo; 2 = algo de acuerdo; 3 = bastante de acuerdo; 4 = totalmente de acuerdo), relativos a la actitud de los adolescentes hacia el profesorado, la policía, las reglas escolares y la ley como sistemas normativos. Las evidencias de validez de constructo

indican que la AAI-A está compuesta por dos factores que en conjunto explican el 39,06% de la varianza. El primer factor fue denominado como Actitud Positiva hacia la Autoridad (APA) y el segundo factor se denomina Actitud Positiva hacia la Transgresión (APT). La fiabilidad por consistencia interna, medida a partir del coeficiente  $\alpha$  de Cronbach, para el primer factor fue de .75, mientras que para el segundo factor fue de .74. Finalmente, en el estudio de Galvez, et al (2015), mediante validación cruzada (Factorial Exploratorio con la primera muestra y análisis confirmatorios con la segunda) se encontró un ajuste adecuado a la estructura bidimensional ( $S-B\chi^2(26) = 116.51$ ,  $p < .001$ ,  $TLI = .90$ ,  $CFI = .93$ ,  $RMSEA = .071$ .), las evidencias de fiabilidad, mediante el clásico  $\alpha$  de Cronbach mostraron valores aceptables ( $APA = .712$ ,  $APT = .756$ ).

#### *Análisis de Datos*

En primer lugar se presentarán los estadísticos descriptivos de centro, dispersión y forma de las 9 variables. Posteriormente se realizarán evaluaciones de las evidencias de validez contrastando los siguientes modelos: M1a Unidimensional (esencialmente) tau-equivalente, fijando las saturaciones factoriales; M1b Unidimensional congénico; M2 Bidimensional correlacionado; M3 Bifactor. Para cada modelo se trabajará con la matriz producto-momento de Pearson, estimando los parámetros con el método de Máxima Verosimilitud Robusta MLM, adecuado para datos no normales (Muthén & Muthén, 1998-2010); además se incorporará métodos para el análisis de datos categóricos mediante el uso de la matriz policórica (Asun et al., 2016; Flora & Curran, 2004; Jöreskog, 1994) utilizando el método de estimación “*Weighted Least Squares Mean and Variance Adjuste*” (mínimos cuadrados ponderados con media y varianza ajustada; WLSMV), adecuado para datos

ordinales (Muthén & Muthén, 1998-2010). Se evaluó la calidad del modelo de reproducir los datos utilizando los siguientes índices de bondad de ajuste; chi-cuadrado de Satorra Bentler ( $SB-\chi^2$ ; Satorra & Bentler, 2001), índice de ajuste comparativo (CFI), índice de Tucker-Lewis (TLI) y error cuadrático medio de aproximación (RMSEA). Para los primeros índices, CFI y TLI, se considera un ajuste adecuado del modelo los valores superiores a .90 (Schumacher & Lomax, 1996) o, con un criterio más exigente, valores superiores a .95 (Schreiber et al., 2006), mientras que para RMSEA se considera un ajuste razonable los valores inferiores a .08 (Browne & Cudeck, 1993) o .06 (Schreiber et al., 2006). Posteriormente se comparó los modelos entre sí, para los métodos basados en matrices de Pearson MLM se estimó la diferencia de  $SB-\chi^2$  utilizando el método de factor de corrección (Asparouhov & Muthén, 2013), por su parte para el modelo ordinal, mediante matrices policóricas, se utilizó la opción DIFFTEST. Estos análisis fueron realizados con Mplus 7.1. Finalmente se evaluaron las evidencias de fiabilidad utilizando los siguientes estimadores:  $\alpha$  de Cronbach,  $\omega$ , GLB y GLBa,  $\alpha$  ordinal y  $\omega$  ordinal. El tamaño muestral fue de 1394 casos.

## Resultados

Los estadísticos descriptivos son presentados en la Tabla 23, se aprecia que, a diferencia del caso anterior, ahora existen ítems con asimetría positiva y otros con asimetría negativa (los valores para el coeficiente estandarizado son superiores a |3|), en lo respecta a la curtosis, se observa que todos los ítems presentan valores negativos (solo 4 ítems presentan valores menores a |3| en el coeficiente estandarizado).

Tabla 23: Estadísticos descriptivos de los ítems del AAI-A

Ítems	Media	Desv. típ.	Asimetría	Error típico	As Estand	Curtosis	Error típico	Cr Estand
ítem 1	3.85	0.963	-0.512	0.066	-7.76	-0.222	0.131	-1.69
ítem 2	3.41	1.181	-0.283	0.066	-4.29	-0.755	0.131	-5.76
ítem 3	2.53	1.265	0.363	0.066	5.50	-0.859	0.131	-6.56
ítem 4	3.29	1.303	-0.231	0.066	-3.50	-1.023	0.131	-7.81
ítem 5	3.99	1.274	-1.019	0.066	-15.44	-0.169	0.131	-1.29
ítem 6	3.35	1.026	-0.272	0.066	-4.12	-0.352	0.131	-2.69
ítem 7	2.22	1.239	0.699	0.066	10.59	-0.568	0.131	-4.34
ítem 8	2.19	1.254	0.738	0.066	11.18	-0.539	0.131	-4.11
ítem 9	2.14	1.287	0.909	0.066	13.77	-0.278	0.131	-2.12

Nota: Desv. típ = Desviación Típica ; As Estand = Coeficiente de asimetría estandarizado; Cr Estand = Coeficiente de curtosis estandarizado

En la Tabla 24 se presentan los índices de ajuste de los modelos, tanto para los basados en correlaciones de Pearson como para los que trabajan con matrices policóricas. Para el caso Unidimensional (M1a y M1b) el ajuste fue del todo inadecuado, independientemente de la matriz a utilizar. Por su parte M2 mostró índices de ajuste bastante satisfactorios, con excepción de un relativamente alto valor de RMSEA en el caso ordinal. De todos los modelos planteados el que mejor ajuste presenta es el Bifactor (M3), mostrando excelentes valores en todos los índices, incluso un valor de  $SB-\chi^2$  no significativo con nivel de alfa  $< .01$  (33,189,  $gl = 18$ ,  $p = .0158$ ) para el caso de matriz de Pearson.



Tabla 24: Ajuste de los modelos para la escala AAIA

Matriz (Método)			Pearson (MLM)					Policóricas (WLSMV)				
Modelo	SB- $\chi^2$	gl	P	CFI	TLI	RMSEA	$\chi^2$	gl	p	CFI	TLI	RMSEA
M1a	2507.811	35	< .001	.054	.027	.225 (.218-.233)	5422,869	35	< .001	.145	.121	.332 (.325-.340)
M1b	828.087	27	< .001	.694	.591	.146 (.137-.155)	1353,496	27	< .001	.790	.719	.188 (.179-.196)
M2	190.927	26	< .001	.941	.918	.065 (.057-.075)	309,391	26	< .001	.955	.938	.088 (.080-.097)
M3	33.189	18	.0158	.994	.988	.025 (.011-.038)	45,794	18	< .001	.996	.991	.033 (.021-.045)

Nota: MLM = Máxima Verosimilitud Robusta; WLSMV = Mínimos Cuadrados ponderados con media y varianza ajustada; M = Modelos; M1a = Modelos Tau-equivalente; M1b = Modelo Congénico; M2: Modelo bidimensional correlacionado; M3 = Modelo Bifactor; SB- $\chi^2$  = Satorra-Bentler Chi Cuadrado; gl = grados de libertad; CFI = índice de ajuste comparativo; TLI = Índice de Tucker-Lewis; RMSEA = Raíz del error cuadrático medio de aproximación;  $\chi^2$  = Chi-cuadrado

Se presenta en la Tabla 25 la comparación de los modelos, para la estimación con MLM basada en correlaciones de Pearson se utilizó el procedimiento mediante el factor de corrección, por su parte para el método WLSMV estimado a partir de la matriz policórica se utilizó la opción DIFFTEST. Se aprecia que al comparar los modelos, de todos ellos el que mejor ajuste presenta es el M3 (al ser la diferencia estadísticamente significativa con M2). Esto indica que el entre un modelo con dos factores y un modelo Bifactor, el que mejor ajuste presenta es el último. Considerando estos análisis, a continuación se presentan los parámetros lambda obtenidos con este modelo así como las evidencias de fiabilidad que se pueden obtener.

Tabla 25: Diferencia entre los modelos mediante el factor de corrección para MLM o mediante DIFFTEST para WLSMV

Pearson MLM	Modelo	$\Delta SB-\chi^2$	$\Delta gl$	p
	M1b-M1a	1.931,986	8	< .001
	M2-M1b	487,8755	1	< .001
	M3-M2	153,3832	8	< .001
Policóricas WLSMV	Modelo	$\Delta SB-\chi^2$	$\Delta gl$	p
	M1b-M1a	3.937,325	8	< .001
	M2-M1b	336.943	1	< .001
	M3-M2	192.564	8	< .001

Nota: MLM = Máxima Verosimilitud Robusta; WLSMV = Mínimos Cuadrados Ponderados con media y varianza ajustada; M1a = Modelos Tau-equivalente; M1b = Modelo Congénico; M2: Modelo bidimensional correlacionado; M3 = Modelo Bifactor.

En la Tabla 26 se muestran las estimaciones de los parámetros lambda obtenidos mediante los métodos MLM y WLSMV (se muestra la representación de la solución para datos categóricos en la Figura 8). Se observa una estructura estable en ambas matrices: Un Factor General (con ítems con saturación negativa 3, 7, 8, 9) y dos factores específicos APA (ítems 1, 4 y 6) y APT (ítems 3, 7, 8 y 9); se menciona que con dos ítems presentaron saturaciones no significativas en el factor APA (pero sí en el Factor General). Se observa además que en casi todos los ítems, los parámetros lambda son mayores para el caso ordinal que para el continuo.

Tabla 26 : Saturaciones con el Modelo Bifactor

Matriz (Método)		Pearson (MLM)			Policórica (WLSMV)		
Factor	FG	APA	APT	FG	APA	APT	
Ítem	Lambda (S.E)	Lambda (S.E)	Lambda (S.E)	Lambda (S.E)	Lambda (S.E)	Lambda (S.E)	
U1	.508 (.051)	.520 (.058)		.580 (.045)	.503 (.055)		
U2	.508 (.031)	<b>.035 (.058)</b>		.515 (.028)	<b>.029 (.048)</b>		
U3	-.309 (.034)		.380 (.031)	-.327 (.031)		.431 (.026)	
U4	.542 (.047)	.484 (.061)		.598 (.042)	.479 (.057)		
U5	.629 (.041)	<b>-.206 (.093)</b>		.720 (.043)	<b>-.324 (.110)</b>		
U6	.586 (.034)	.318 (.056)		.642 (.029)	.290 (.051)		
U7	-.213 (.035)		.654 (.030)	-.257(.033)		.698 (.022)	
U8	-.381 (.034)		.770 (.028)	-.428 (.031)		.804 (.023)	
U9	-.467 (.033)		.488 (.032)	-.504 (.030)		.537 (.025)	

Nota: En negritas los parámetros lambda no significativos.

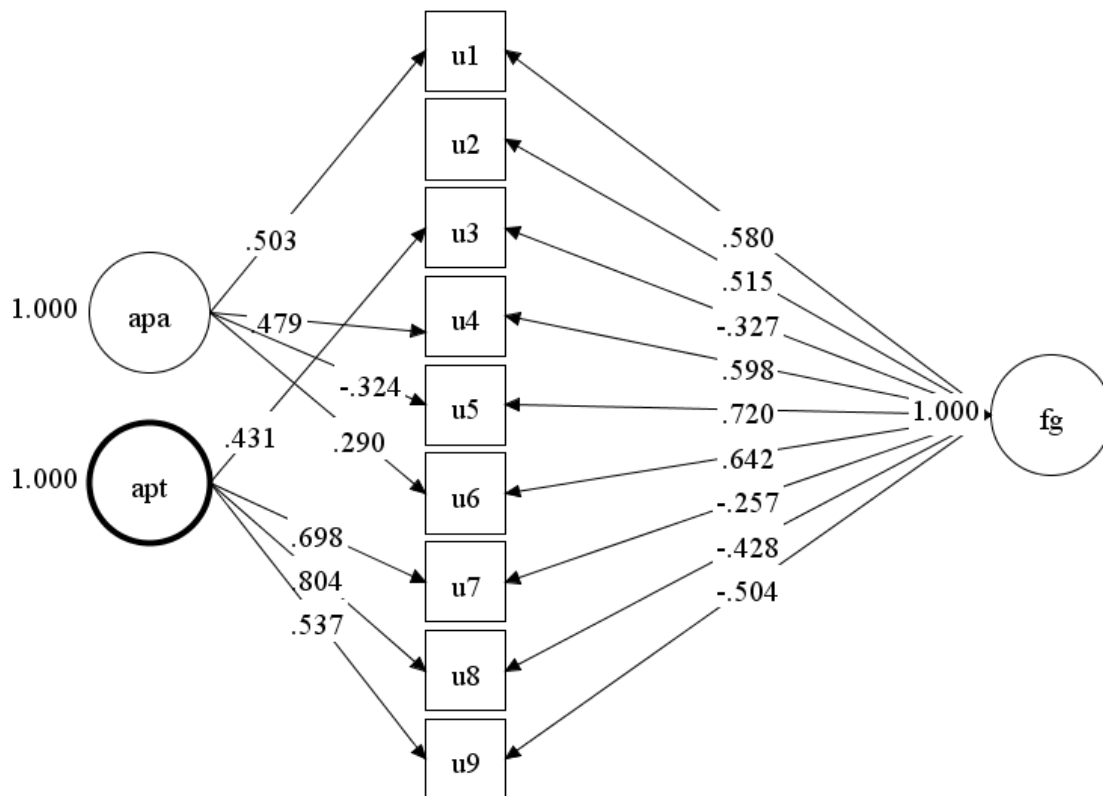


Figura 8: Gráfico del modelo Bifactor con datos categóricos para AAI-A.

### *Evidencias de fiabilidad*

Una vez evaluado el modelo factorial que mejor se ajusta a los datos se procede a obtener evidencias de fiabilidad. En la Tabla 27 se presenta la estimación de fiabilidad obtenida para M3 (recodificando los ítems negativos). Se observa que para el FG los coeficientes que asumen unidimensionalidad (con excepción de GLBa), presentan sobrestimaciones de la fiabilidad, tal como señala la literatura y los estudios de simulación previos, por ejemplo  $\alpha$  muestra un valor de .77, en tanto  $\omega_t$  (en su versión lineal) y GLB coinciden en un valor de .83, sin embargo la fiabilidad estimada con  $\omega_h$  presenta un valor mucho más realista en torno a .60.

Tabla 27: Evidencias de fiabilidad para AAI-A basado en M3

	$\alpha$	$\omega_t$	$\omega_s$	$\omega_h$	GLB	GLBa	$\alpha$ ordinal	$\omega_t$ ordinal	$\omega_s$ ordinal	$\omega_h$ ordinal
APA	.70	.763	.112	-	.775	.750	.74	.831	.183	-
APT	.76	.773	.57	-	.622	.543	.81	.825	.599	-
FG	.77	.83	-	.599	.83	.66	.81	.87	-	.62

Nota:  $\alpha$  = coeficiente alfa;  $\omega_t$  = coeficiente omega total;  $\omega_s$  = coeficiente omega subescala;  $\omega_h$  = coeficiente omega jerárquico; GLB = Greatest Lower Bound (glb.fa); GLBa = Greatest Lower Bound (glba.algebraic);  $\alpha$  ordinal = coeficiente alfa ordinal;  $\omega_t$  ordinal = coeficiente omega total ordinal;  $\omega_s$  ordinal = coeficiente omega subescala ordinal;  $\omega_h$  = coeficiente omega jerárquico ordinal

### **Conclusión**

Considerando la estructura multidimensional subyacente, con un factor general común a todos los ítems y dos factores grupales, la estimación más adecuada la entregaría el coeficiente  $\omega_h$ , debido a que los coeficientes que asumen unidimensionalidad estarían entregando sobrestimaciones de la fiabilidad, con valores artificialmente altos que no reflejan la fiabilidad de las puntuaciones respecto al factor general, tal como se mostró en este

segundo ejemplo empírico. Además, dependiendo de la matriz de correlación a analizar, estimaciones lineales u ordinales, el valor de la fiabilidad oscilará entre .60 con Pearson y .62 con policóricas. Considerando que son solo 4 categorías, parece más prudente trabajar con el modelo ordinal, pues se evitan los sesgos que produce la categorización en la recuperación de parámetros lambda, especialmente cuando las variables son asimétricas. La única excepción a lo esperado es el coeficiente GLBa el cual presenta resultados cercanos a los de  $\omega_h$ .



## DISCUSIÓN Y CONCLUSIONES FINALES.

### 1. ESTIMACIÓN DE LA FIABILIDAD EN MODELOS UNIDIMENSIONALES

Para el caso de datos continuos cuatro factores fueron manipulados: Modelo Tau-equivalente y Congenérico, tamaño muestral (250, 500 y 1000), el tamaño del test (6 y 12 ítems) y el número de ítems asimétricos (desde 0 ítems asimétricos hasta el total de ítems asimétricos), con el objetivo de evaluar la robustez a la presencia de asimetría de los cuatro coeficientes de fiabilidad analizados. Los resultados son discutidos a continuación bajo el supuesto de unidimensionalidad.

En condiciones de tau-equivalencia, los coeficientes  $\alpha$  y  $\omega$  convergen, sin embargo, en ausencia de tau-equivalencia,  $\omega$  presenta siempre mejores estimaciones, menor RMSE y sesgo que  $\alpha$ . Por lo que, en esta condición más realista (Green & Yang, 2009a; Yang & Green, 2011),  $\alpha$  se vuelve un estimador negativamente sesgado de la fiabilidad (Cho & Kim, 2015; Graham, 2006; Sijtsma, 2009a) y  $\omega$  es siempre preferible a  $\alpha$  como estimador de la fiabilidad (Dunn et al., 2014). En el caso de no violación del supuesto de normalidad,  $\omega$  es el mejor estimador de todos los coeficientes evaluados (Revelle & Zinbarg, 2009).

En relación con el tamaño muestral, se observa que este factor tiene un efecto pequeño en los coeficientes  $\alpha$  y  $\omega$ , mejorando ligeramente su estimación reduciendo el RMSE y el sesgo a medida que aumenta el tamaño muestral, sin embargo se puede afirmar que para estos dos coeficientes, con tamaño muestrales de 250 continuidad y normalidad se obtienen estimaciones relativamente precisas (Javali, Gudaganavar, & Raj, 2011; Tang & Cui, 2009),

siempre y cuando se trabaje con un modelo unidimensional. Para los coeficientes GLB y GLBa, a medida que aumenta el tamaño muestral los valores de RMSE y porcentaje de sesgo tienden a disminuir; sin embargo, mantienen el sesgo positivo para la condición de normalidad incluso con muestras grandes de tamaño 1000 (Shapiro & ten Berge, 2000; Sijtsma, 2009a; ten Berge & Sočan, 2004).

Con relación al tamaño del test, se observa mayores valores de RMSE y sesgo para la condición de 6 ítems que para la de 12, por lo que se entiende que a mayor número de ítems menor es el RMSE y sesgo de los estimadores (Cortina, 1993).

En lo que respecta al efecto de incorporar progresivamente ítems asimétricos al conjunto de datos, se observa que el coeficiente  $\alpha$  es altamente sensible a ítems asimétricos, resultados similares a los encontrados por Sheng y Sheng (2012) y Green y Yang (2009b). El coeficiente  $\omega$  presenta valores similares de RMSE y sesgo a los de  $\alpha$ , pero ligeramente mejores, incluso en presencia de un modelo tau-equivalente. Por su parte, GLB presenta mejores estimaciones cuando la asimetría media de los ítems ronda valores de .30, compensando su sesgo positivo. De este modo, cuando la asimetría media de los ítems aumenta, GLB presenta mejores estimaciones con tamaños muestrales pequeños que con tamaños grandes, a similar conclusión se llega con GLBa que presenta mejores estimaciones que  $\omega$  con una asimetría media en torno a valores de .20, sin embargo a medida que aumenta la asimetría, valores de .70, GLB muestra un mejor desempeño que GLBa.



Considerando que en la práctica es común encontrarse con datos asimétricos (Ho & Yu, 2015; Micceri, 1989; Norton et al., 2013), la sugerencia de Sijtsma (2009a) de usar GLB como estimador de la fiabilidad se muestra como acertada. Otros autores como Revelle y Zinbarg (2009) recomiendan el uso de  $\omega$  (Crutzen & Peters, 2015; Dunn et al., 2014; Peters, 2014; Teo & Fan, 2013). Sin embargo, este coeficiente solo produce buenos resultados en condiciones de normalidad o con poca asimetría. En cualquier caso, estos coeficientes presentan grandes ventajas teóricas y empíricas frente a  $\alpha$ .

Respecto a la caso de datos categóricos y dicotómicos el objetivo de esa simulación era comparar las estimaciones realizadas con cuatro coeficientes de fiabilidad en presencia de datos categorizados y asimétricos, manteniendo el supuesto de unidimensionalidad en un modelo congénico (ítems con diferente discriminación, o lo que es equivalente, con diferente saturación factorial estimada). Los resultados muestran que los coeficientes lineales  $\alpha$  y  $\omega$ , estimados a partir de la matriz de correlación de Pearson, van presentando un peor rendimiento (lo que se traduce en mayores infraestimaciones y más alto RMSE) a medida que disminuye el número de categorías de respuesta (Gadermann et al., 2012; Lozano et al., 2008; Zumbo et al., 2007), así también se ha observado la presencia de sesgos negativos a medida que aumenta la asimetría de la distribución del test, esto es compatible con la literatura previa la cual señala que la asimetría y la categorización tienen un efecto negativo en la estimación de los parámetros de discriminación producida al trabajar con matrices de correlación producto-momento de Pearson (Flora et al., 2012; Holgado-Tello et al., 2010). Solo cuando existen 5 opciones de respuesta y distribuciones normales la infraestimación generada por  $\alpha$  y  $\omega$  no es demasiado grave. En contraste, la capacidad de recuperar el

parámetro de fiabilidad simulado por parte de las versiones ordinales de estos coeficientes, utilizando matrices de correlación tetracórica y policórica supera ampliamente a sus versiones lineales en todas las condiciones simuladas, mostrándose como alternativas adecuadas en condiciones de datos categóricos (Elosua & Zumbo, 2008; Gadermann et al., 2012; Zumbo et al., 2007), incluso en condiciones con poca asimetría. Sin embargo, independientemente de la matriz seleccionada para su estimación, se observa como el coeficiente  $\alpha$  produce una infestimación constante al estar trabajando bajo un modelo congénico (Graham, 2006; Lord & Novick, 1968; Raykov, 1997, 2001). Así el coeficiente  $\omega$  ordinal presenta mejor desempeño que  $\alpha$  ordinal.

En tanto, para los coeficientes GLB se observa en general un mejor funcionamiento que las versiones lineales de  $\alpha$  y  $\omega$ , mostrando una relativamente buena capacidad de recuperación en ciertas condiciones excepto para el caso dicotómico, donde presenta sesgos negativos altos en todas las condiciones; por su parte el procedimiento GLBa mostró menor capacidad de recuperación de la fiabilidad, presentando valores de sesgo por encima de .10 en las condiciones de tres o menos categorías de respuesta. Conviene recordar que los resultados aquí presentados y discutidos han sido obtenidos con modelos unidimensionales, sin embargo en la realidad este tipo de estructuras tiende a limitarse a constructos psicológicos restringidos “*narrow constructs*” (Reise, 2012; Reise et al., 2000), en consecuencia estructuras más complejas que evalúen multidimensionales deben ser abordadas.

## 2. ESTIMACIÓN DE LA FIABILIDAD EN MODELOS MULTIDIMENSIONALES.

En situaciones donde los datos responden a estructuras multidimensionales o con errores positivamente correlacionados, condiciones que hemos tratado conjuntamente, ya que la correlación entre errores como hemos ejemplificado tiende a producir estructuras complejas, la estimación de la fiabilidad por parte de coeficientes que asumen el supuesto de unidimensionalidad puede generar sobreestimaciones de la fiabilidad atribuida al factor general (Green & Yang, 2015; Raykov, 2001; Yang & Green, 2011; Zinbarg et al., 2005, 2006), siendo por lo tanto más precisas las estimaciones ofrecidas por el coeficiente  $\omega_h$  (Green & Yang, 2015; Reise, 2012; Revelle & Zinbarg, 2009; Zinbarg et al., 2005, 2006).

Así también, el estudio de monte Carlo mosró que en presencia de ítems redundantes en condiciones donde la fiabilidad del factor general es cercana a 0 (por lo que en vez de tener una estructura claramente bifactor se obtiene en realidad una estructura con tres factores escasamente correlacionados entre sí) los coeficientes de fiabilidad que asumen unidimensionalidad muestran altos valores, sin embargo tal como señala Cortina (1993) ¿Cuál es la interpretación de esa puntuación global de fiabilidad si no existe una estructura que la sustente? En este sentido, si lo que interesa al investigador es evaluar la proporción de varianza del test explicada por todos los factores, el coeficiente  $\alpha$  genera el menor valor, en tanto los coeficientes  $\omega_t$  y GLB generan practicamente los mismos resultados, al trabajar con muestral de tamaño 1000, de todos ellos GLBa obtiene el mayor valor. De este modo se contribuye al debate en torno a cuál de los dos estimadores,  $\omega_t$  (Revelle & Zinbarg, 2009) o GLB (Sijtsma, 2009a, 2009b) es el mejor estimador de la varianza atribuida a todos los factores, ambos producen resultados casi identicos (al evaluar  $\omega_t$  dentro de un modelo

Bifactor). De hecho, en el segundo estudio empírico con datos reales, el valor de estos dos coeficientes resultó de .83 (trabajando con la versión lineal de  $\omega_t$ ).

Si lo que interesa es evaluar la fiabilidad atribuible al factor general, común a los ítems, el coeficiente adecuado será  $\omega_h$ . Los estudios de simulación mostraron que los coeficientes que asumen unidimensionalidad presentan sesgos positivos elevados al ser utilizados en contextos multidimensionales o con errores positivamente correlacionados (Cortina, 1993; Raykov, 2001; Revelle & Zinbarg, 2009; Zinbarg et al., 2005, 2006). De este modo, las precauciones de que fiabilidades por encima de un determinado valor, por ejemplo .90 (Streiner, 2003) evitan problemas de redundancia son discutibles, ya que se ha visto que tanto el coeficiente  $\alpha$  como los demás estimadores unidimensionales, pueden inflar la fiabilidad mostrando valores de, por ejemplo, .80, cuando la fiabilidad atribuible al factor general es significativamente menor. En consecuencia, la solución no es poner un punto de corte arbitrario al valor de la fiabilidad, sino utilizar un procedimiento adecuado que evite el sesgo debido a la redundancia como el coeficiente  $\omega_h$ , para lo cual es necesario evaluar previamente la estructura subyacente a los datos (Crutzen & Peters, 2015; Green & Yang, 2009a, 2015; Revelle & Zinbarg, 2009) específicamente mediante el modelo Bifactor (Reise, 2012; Rios & Wells, 2014).

Del mismo modo, resulta una práctica habitual dentro de SEM evaluar el ajuste de estructuras en las cuales se ha incorporado una correlación entre los pares de dos ítems (por ejemplo en presencia de ítems con solapamiento de contenido). Si el investigador ha decidido, tanto por razones sustantivas como por razones de mejorar el ajuste, incorporar la correlación entre errores, debe del mismo modo corregir la estimación de la fiabilidad controlando la covarianza entre errores mediante el procedimiento descrito por Raykov

(2001). En caso de omitir esa corrección, las estimaciones de la fiabilidad con procedimientos unidimensionales (como el coeficiente  $\alpha$ ) entregarán valor inflados de la fiabilidad del factor general (Cho & Kim, 2015; Zinbarg et al., 2005, 2006).

Con respecto al coeficiente  $\alpha$ , no solo es discutible que sea un límite inferior de la fiabilidad ante el incumplimiento de sus supuestos, sino que en ciertas condiciones, ante violaciones del supuesto de unidimensionalidad presenta severas sobreestimaciones, entregando valores inflados al no diferenciar varianza del factor general de la varianza de los factos específicos (o de la correlación entre errores).

En la práctica resulta frecuente encontrarse con datos en los que se incumpla en algún grado el supuesto de unidimensionalidad (Green & Yang, 2015; ten Berge & Sočan, 2004), por lo es imperativo evaluar las evidencias de validez, es decir la estructura subyacente al conjunto de ítems, así como los supuestos del coeficiente  $\alpha$  previo a realizar análisis de fiabilidad (Cho & Kim, 2015; Crutzen & Peters, 2015; Green & Yang, 2009a). Así los ejemplos empíricos aquí analizados muestran que el coeficiente  $\alpha$  tiende a generar estimaciones con potenciales sesgos de infraestimación (para casos unidimensionales congéntricos) o sobrestimación (en contextos multidimensionales bajo modelo Bifactor o con correlación positiva entre los errores de los ítems). De todos los coeficientes evaluados en los estudios con datos reales, el único que presentó una anomalía fue el coeficiente GLBa que por las correlaciones negativas entre los ítems (incluso invirtiendo algunos de ellos) presenta valores muy cercanos al del coeficiente  $\omega_h$ .

### 3 CONCLUSIONES GENERALES

En síntesis el coeficiente  $\alpha$  tiende a infraestimar la fiabilidad cuando se trabaja con datos congéntricos (Cronbach, 1951; Green & Yang, 2009a; Lord & Novick, 1968; Sijtsma, 2009a; Yang & Green, 2011), también se ve afectado negativamente frente a asimetría en los ítems (Sheng & Sheng, 2012), generando infraestimaciones cuando se utiliza con ítems categóricos (Elosua & Zumbo, 2008; Gadermann et al., 2012; Green & Yang, 2009b; Zumbo et al., 2007). En estas condiciones se puede argumentar que el coeficiente  $\alpha$  se vuelve un “límite inferior de la fiabilidad”, sin embargo como señala Raykov (2001) en presencia de errores positivamente correlacionados o multidimensionalidad  $\alpha$  tiende a generar sobreestimaciones de la fiabilidad, indicando que algunas medidas son fiables cuando en realidad no lo son, de hecho puede ofrecer altos valores de fiabilidad en condiciones donde la fiabilidad es cercana a 0. En consecuencia, este coeficiente se convierte en un estimador que puede estar sesgado negativa o positivamente, el sesgo final dependerá de la compleja interacción entre el grado de cumplimiento (o incumplimiento) de los supuestos mencionados (Yang & Green, 2011), convirtiendo al coeficiente  $\alpha$  en un estimador inestable de la fiabilidad.

En la literatura psicométrica existe amplio consenso entre los teóricos de lo inadecuado e inapropiado del coeficiente  $\alpha$  en condiciones realistas (Deng & Chan, 2016; Dunn et al., 2014; Green & Yang, 2009b, 2009a; Peters, 2014; Revelle & Zinbarg, 2009; Sijtsma, 2009a; Sijtsma & van der Ark, 2015; Teo & Fan, 2013; Yang & Green, 2011) invitando a los investigadores aplicados a abandonar el uso de  $\alpha$  por el de otros coeficientes más adecuados. Es aquí donde no existe pleno consenso, ya que diferentes autores han propuesto distintas

alternativas para superar las limitaciones que presenta  $\alpha$ . En este trabajo se optó por evaluar dos alternativas a  $\alpha$ : El coeficiente  $\omega$  (McDonald, 1999; Revelle & Zinbarg, 2009; Zinbarg et al., 2005) y el coeficiente GLB (Sijtsma, 2009a; ten Berge & Sočan, 2004).

En el estudio 1 de esta tesis se presentó el efecto que la asimetría tiene en la estimación de la fiabilidad con los coeficientes  $\alpha$ ,  $\omega$  y dos procedimientos para obtener GLB (glb.fa y glb.algebraic). Se mostró que  $\omega$  era relativamente más robusto que  $\alpha$  en condiciones de asimetría baja del test (pocos o ningún ítem asimétrico), sin embargo a medida que se incorporaban ítems con asimetría las estimaciones de  $\alpha$  y  $\omega$  iban empeorando, en este sentido, GLB mostró una mayor robustez al recuperar al parámetro simulado con menor sesgo que los otros dos coeficientes. Además, considerando que en la práctica habitualmente se trabaja con escalas tipo Likert o ítems dicotómicos, se evaluó el efecto que ítems categóricos y asimétricos tienen en la estimación de la fiabilidad, mostrando que se producen sesgos en la estimación de los parámetros  $\lambda$  los cuales van aumentando en la medida en que existan menos categorías de respuesta y los ítems sean más asimétricos, en consecuencia se muestra que las versiones ordinales de los coeficientes  $\alpha$  y  $\omega$  producen una mejor recuperación del parámetro simulado. Sin embargo en este estudio solo se evaluó una estructura unidimensional. En este sentido, el segundo estudio muestra el efecto que la ausencia de unidimensionalidad (tanto por la presencia de ítems redundantes, que genera correlaciones positivas entre los ítems, como en estructuras Bifactor propiamente tal) tiene en la estimación de la fiabilidad mediante coeficientes que asumen unidimensionalidad, generando groseras sobrestimaciones, mostrando por ejemplo valores altos de fiabilidad donde la fiabilidad del factor general es mucho menor, por lo que se deben realizar las correcciones adecuadas al estimar la fiabilidad, por ejemplo si existe correlación entre los

errores de un par de ítems (práctica habitual para mejorar el ajuste de los modelos) se debe incorporar la covariazación entre esos elementos en la estimación de la fiabilidad, en caso contrario los coeficientes estarán inflados. Finalmente, y para ejemplificar, se analizan las evidencias de validez y de fiabilidad de dos instrumentos aplicados a una muestra chilena, mostrando en general resultados concordantes con lo esperado, considerando tanto la literatura como los estudios de simulación previos mostrados en la tesis.

Por lo tanto, la familia de coeficientes  $\omega$ , cada uno apropiado para situaciones distintas, ha mostrado recuperar de forma más adecuada la fiabilidad que el coeficiente  $\alpha$ , por ejemplo en el caso de datos congenéricos y unidimensionales, el coeficiente  $\omega_t$  no se ve afectado negativamente, a diferencia de  $\alpha$ . Además el coeficiente  $\omega$  ordinal permite recuperar adecuadamente la fiabilidad simulada en condiciones de datos congenéricos y asimétricos, de dos a cinco categorías de respuesta. En relación a condiciones multidimensionales, como se ha visto en el Estudio 2, el coeficiente  $\alpha$  produce sesgos positivos incluso cuando existe un solo ítem redundante (es decir, con correlación positiva entre errores de dos ítem), incrementando artificialmente la fiabilidad de la puntuación del test para el factor general. En estas condiciones el coeficiente  $\omega_t$  genera una mayor sobrestimación que  $\alpha$ . Por lo que, si se evalúa un modelo Bifactor, el estimador más adecuado de la fiabilidad del factor general sería  $\omega_h$ , ya que no comete los graves sesgos positivos por los coeficientes que asumen unidimensionalidad controlando la varianza atribuida a los factores grupales.

Por su parte el coeficiente GLB ha mostrado en modelos unidimensionales y con asimetría, recuperar mejor la fiabilidad que los coeficientes  $\alpha$  y  $\omega$ , esto es probablemente debido a una compensación de sesgos entre la ligera sobrestimación que presenta en



condiciones de normalidad y el efecto negativo de la asimetría en la estimación de estos coeficientes, generando por lo tanto en condiciones específicas resultados adecuados mejores que los de  $\alpha$  e incluso que los de  $\omega$ . Del mismo modo, en condiciones de datos categóricos presenta un mejor funcionamiento que los estimadores lineales de  $\alpha$  y  $\omega$ . Por otro lado, en condiciones de multidimensionalidad, el coeficiente GLB mediante el procedimiento glb.fa, genera resultados casi idénticos a los de  $\omega_i$  en la misma condición, sobrestimando naturalmente la fiabilidad del factor general, del mismo modo, el procedimiento glb.algebraic genera sobrestimaciones incluso mayores que las de glb.fa. En este sentido, estos procedimientos son muy sensibles a la presencia de covariación entre errores, considerando su estimación siempre asintótica y bajo el supuesto de unidimensionalidad.

Dentro de las limitaciones de esta tesis se pueden mencionar que las evaluaciones de la fiabilidad no consideraron estructuras multinivel (Geldhof, Preacher, & Zyphur, 2014), por lo que se recomienda desarrollar líneas de investigación en el futuro que trabajen sobre estructuras con más de un nivel. Otra limitación importante es que no se desarrollaron ni se ampliaron las relaciones entre Teoría de Respuesta al Ítem y el Análisis Factorial de Ítems, especialmente en lo relativo a la fiabilidad de las puntuaciones, donde en el primer marco conceptual se evalúa mediante la Función de Información que tiene la ventaja de no asumir igualdad para todos los niveles o rangos de la variable latente, en este sentido McDonald (1999) señala que el coeficiente  $\omega$  es equivalente a una media de la Función de Información. En esta tesis tampoco se evaluó la interacción entre diferentes valores de asimetría (presencia en un mismo test de ítems con asimetría positiva e ítems con asimetría negativa), seleccionando solamente valores en torno a -1. Del mismo modo, en este estudio no se presentaron modelos con errores correlacionados negativamente. En relación al caso ordinal,

cabe mencionar que Green y Yang (Green & Yang, 2009b) han desarrollado un coeficiente no lineal alternativo que incorpora los parámetros umbrales o tau para estimar la fiabilidad, así considera la naturaleza de la escala pero, a diferencia de las versiones ordinales de  $\alpha$  y  $\omega$ , no estima la fiabilidad de la variable subyacente  $X^*$ , abriendo el debate a la diferencia entre fiabilidad simulada y fiabilidad empírica (obtenida como la suma de los ítems categorizados) sin embargo este coeficiente no fue evaluado en esta tesis, siendo interesante realizar futuros estudios que incorporen el coeficiente no lineal de Green y Yang. Del mismo modo, existen otros método de estimación de la fiabilidad para las puntuaciones factoriales, las cuales se puede obtener con el procedimiento “*Bayes expected a posteriori*”, el nuevo coeficiente fue denominado como “*Overall Reliability of fullyInformative prior Oblique N-EAP scores (ORION)*” y considera además la matriz de correlación entre los factores para soluciones oblicuas (Ferrando & Lorenzo-seva, 2016), resulta por lo tanto interesante evaluar el funcionamiento de estos coeficientes como una futura línea de investigación. Otra limitación tiene que ver con que al estimar los parámetros del modelo multidimensional no se asumieron posibles errores de especificación del modelo, de esta forma el modelo factorial muestral estuvo siempre adecuadamente especificado (considerando el modelo poblacional), en la práctica sin embargo es posible que existan errores cuando se especifican los modelos debido a que normalmente el modelo poblacional subyacente a un conjunto de datos resulta, en un principio, desconocido. Se debe considerar en futuros estudios el efecto que tiene en la estimación de la fiabilidad (y por consiguiente de los parámetros de discriminación lambda) la presencia de posibles errores de especificación.

## REFERENCIAS

- Abad, F. J., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid: Editorial Síntesis.
- Al-Homidan, S. (2008). Semidefinite programming for the educational testing problem. *Central European Journal of Operations Research*, 16(3), 239–249. doi:10.1007/s10100-008-0056-0
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *The Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Asun, R. A., Rdz-Navarro, K., & Alvarado, J. M. (2016). Developing multidimensional likert scales using item factor analysis : The case of four-point items. *Sociological Methods & Research*, 45(1), 109–133. doi:10.1177/0049124114566716
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data : A demonstration using FACTOR. *Practical Assessment, Research & Evaluation*, 19(5), 1–14. doi:10.1146/annurev.psych.53.100901.135239
- Bernaards, C., & Jennrich, R. (2015). Package “GPArotation.” Retrieved from <http://ftp.daum.net/CRAN/web/packages/GPArotation/GPArotation.pdf>
- Bollen, K. A. (2002). Latent variable in psychology and social sciences. *Annual Review of Psychology*, 53, 605–634.

- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. (Guildford, Ed.). New York: Guildford.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment. Quantitative Applications in the Social Sciences* (Vol. 17). Sage publication Inc. Retrieved from [http://books.google.com/books?id=BN\\_MMD9BHogC](http://books.google.com/books?id=BN_MMD9BHogC)
- Cava, M. J., Estévez, E., Buelga, S., & Musitu, G. (2013). Propiedades psicométricas de la Escala de Actitudes hacia la Autoridad Institucional en adolescentes (AAI-A). *Anales de Psicología*, 29(2), 540–548. doi:10.6018/analesps.29.2.138031
- Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods*, 18(2), 207–230. doi:10.1177/1094428114555994
- Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. Retrieved from <http://psycnet.apa.org/journals/apl/78/1/98/>
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. doi:10.1007/BF02310555
- Crutzen, R., & Peters, G.-J. Y. (2015). Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review*, 1–14. doi:10.1080/17437199.2015.1124240
- Deng, L., & Chan, W. (2016). Testing the difference between reliability coefficients alpha and omega. *Educational and Psychological Measurement, online*, 1–19. doi:10.1177/0013164416658325

- Distefano, C. (2002). A the impact of categorization with confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 327–346. doi:10.1207/S15328007SEM0903
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. doi:10.1111/bjop.12046
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15(2), 315–321.
- Elosua, P., & Zumbo, B. D. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20(4), 896–901.
- Ferrando, P. J., & Lorenzo-seva, U. (2013). Unrestricted item factor analysis and some relations with Item Response Theory. Department of Psychology, Universitat Rovira i Virgili, Tarragona.
- Ferrando, P. J., & Lorenzo-seva, U. (2016). A note on improving EAP trait estimation in oblique factor-analytic and item response theory models. *Psicológica*, 37, 235–247.
- Ferrando, P. J., & Lorenzo-Seva, U. (2014). El análisis factorial exploratorio de los ítems: algunas consideraciones adicionales. *Anales de Psicología*, 30(3), 1170–1175. doi:10.6018/analesps.30.3.199991
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466–491. doi:10.1037/1082-989X.9.4.466
- Flora, D. B., LaBrish, C., & Chalmers, R. P. (2012). Old and new ideas for data screening

- and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Psychology*, 3(MAR), 1–21. doi:10.3389/fpsyg.2012.00055
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. doi:10.1017/CBO9781107415324.004
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data : A conceptual , empirical , and practical guide. *Practical Assessment, Research & Evaluation*, 17(3).
- Gálvez-Nieto, J.L., Vera-Bachman, V., Trizano, I., & García, J. A. (2015). Examen psicométrico de la escala de Actitudes hacia la Autoridad Institucional (AAI-A), en estudiantes chilenos. *Revista Iberoamericana de Diagnóstico Y Evaluación*, 1(39), 57–67.
- Gálvez-Nieto, J. L., Vera-Bachman, D., & Trizano, I. (2015). Estudio confirmatorio del cuestionario para evaluar clima social del centro escolar en Chile. *Revista Mexicana de Psicología*, 32(2), 160–168.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72–91. doi:10.1037/a0032138
- Gessaroli, M., & Folske, J. (2002). Generalizing the reliability of tests comprised of testlets. *International Journal of Testing*, 2(3), 277–295. doi:10.1080/15305058.2002.9669496
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability what they are and how to use them. *Educational and Psychological*

- Measurement*, 66(6), 930–944. doi:10.1177/0013164406288165
- Green, S. B., & Yang, Y. (2009a). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74(1), 121–135. doi:10.1007/s11336-008-9098-4
- Green, S. B., & Yang, Y. (2009b). Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha. *Psychometrika*, 74(1), 155–167. doi:10.1007/s11336-008-9099-3
- Green, S. B., & Yang, Y. (2015). Evaluation of dimensionality in the assessment of internal consistency reliability: Coefficient alpha and omega coefficients. *Educational Measurement: Issues and Practice*, 34(4), 14–20. doi:10.1111/emip.12100
- Gulliksen, H. (1950). *Theory of Mental Tests*. Hillsdale, NJ: Lawrence Erlbaum.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255–282. doi:10.1007/BF02288892
- Headrick, T. C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics and Data Analysis*, 40(4), 685–711. doi:10.1016/S0167-9473(02)00072-5
- Ho, A. D., & Yu, C. C. (2015). Descriptive statistics for modern test score distributions: skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement*, 75(3), 365–388. doi:10.1177/0013164414548576
- Hogan, T. P., Benjami, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, 60(523–531).

- Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity*, 44(1), 153–166. doi:10.1007/s11135-008-9190-y
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling an overview and a meta-analysis. *Sociological Methodology*, 26(3), 329–367. doi:10.1177/0049124198026003003
- Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, 6(3), 153–160. doi:10.1007/BF02289270
- Hunt, T. D., & Bentler, P. M. (2015). Quantile lower bounds to population reliability based on locally optimal splits. *Psychometrika*, 80(1), 182–195. doi:10.1007/s11336-013-9393-6.Quantile
- Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika*, 42(4), 567–578. doi:10.1007/BF02295979
- Javali, S. B., Gudaganavar, N. V., & Raj, S. M. (2011). Effect of Varying Sample Size in Estimation of Coefficients of Internal Consistency. Retrieved from [https://www.webmedcentral.com/wmcpdf/Article\\_WMC001649.pdf](https://www.webmedcentral.com/wmcpdf/Article_WMC001649.pdf)
- Jennrich, R., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, 76(4), 537–549. doi:10.1007/s11336-011-9218-4.
- Jöreskog, K. G. (1994). On the estimation of polychorics correlations and their asymptotic covariance matrix. *Psychometrika*, 59(3), 381–389. doi:10.1007/BF02296131



- Jöreskog, K. G., & Sorbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge: MA: Abt Books.
- Jöreskog, K. G., & Sorbom, D. (1996). *LISREL 8: User's reference guide*. Scientific Software International.
- Kane, M. (2008). *Errors of Measurement, Theory, and Public Policy*. Policy.
- Kubinger, K. D. (2003). On artificial results due to using factor analysis for dichotomous variables. *Psychology*, 45(1), 106–110. Retrieved from [http://www.pabst-publishers.de/psychology-science/1-2003/pdf\\_06.pdf](http://www.pabst-publishers.de/psychology-science/1-2003/pdf_06.pdf)
- Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika*, 2, 151–160.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: what did they really say? *Organizational Research Methods*, 9(2), 202–220. doi:10.1177/1094428105284919
- Lee, G., & Frisbie, D. A. (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets. *Applied Measurement in Education*, 12(3), 237–255. doi:10.1207/S15324818AME1203\_2
- Lila, M., Oliver, A., Catalá-Miñana, A., Galiana, L., & Gracia, E. (2014). The Intimate Partner Violence Responsibility Attribution Scale (IPVRAS). *The European Journal of Psychology Applied to Legal Context*, 6(1), 29–36.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Menlo Park, CA: Addison-Wesley Publishing Company.

- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, Instruments and Computers*, 38(1), 88–91.
- Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2: A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and irt models. *Applied Psychological Measurement*, 37(6), 497–498. doi:10.1177/0146621613487794
- Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 4(2), 73–79. doi:10.1027/1614-2241.4.2.73
- Lumsden, J. (1976). Test theory. *Annual Review of Psychology*, 27(1), 251–280. doi:10.1146/annurev.ps.22.020171.001205
- Martínez-Arias, M. R., Hernández, M. J., & Hernández, M. V. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Martínez-Arias, R. (1995). *Psicometria: Teoría de los Test psicológicos y educativos*. Madrid: Editorial Síntesis.
- Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, 12(2), 157–176. doi:10.1037/1082-989X.12.4.433
- Maydeu-Olivares, A., Kramp, U., García-Forero, C., Gallardo-Pujol, D., & Coffman, D. (2009). The effect of varying the number of response alternatives in rating scales: Experimental evidence from intra-individual effects. *Behavior Research Methods*,

41(2), 295–308. doi:10.3758/BRM.41.2.295

McDonald, R. (1999). *Test Theory: a unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156–166. doi:10.1037/0033-2909.105.1.156

Moltner, A., & Revelle, W. (2015). Find the greatest lower bound to reliability. Retrieved from <http://personality-project.org/r/psych/help/glb.algebraic.html>

Muñiz, J. (1998). La medición de lo psicológico. *Psicothema*, 10(1), 1–21.

Muñiz, J. (2000). *Teoría Clásica de los Test*. Madrid: Pirámide.

Muñiz, J., & Fonseca-Pedrero, E. (2008). Construcción de Instrumentos de Medida para la Evaluación Universitaria. *Revista de Investigación En Educación*, 5, 13–25.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115–132. doi:10.1007/BF02294210

Muthén, L. K., & Muthén, B. O. (n.d.). *Mplus User's Guide* (Sixth Edit). Los Ángeles, CA: Muthén & Muthén.

Norton, S., Cosco, T., Doyle, F., Done, J., & Sacker, A. (2013). The Hospital Anxiety and Depression Scale: A meta confirmatory factor analysis. *Journal of Psychosomatic Research*, 74(1), 74–81. doi:10.1016/j.jpsychores.2012.10.010

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1–13.

- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Nunnally, J. C., & Bernstein, I. (1995). *Teoría Psicométrica* (3rd ed.). México: McGraw-jill/Interamericana de México, S.A. de C. V.
- Pearson, K. (1900). Mathematical Contributions to the Theory of Evolution. VII. On the Correlation of Characters not Quantitatively Measurable. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 195(1900), 1–405. doi:10.1098/rsta.1900.0022
- Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity. *The European Health Psychologist*, 16(2), 56–69.
- Prieto, G., & Delgado, A. (2010). Fiabilidad y Validez. *Papeles Del Psicólogo*, 31(1), 67–74.
- Racine, J. (2012). Rstudio: A plataform-independet IDE for R and Sweave. *Journal of Applied Econometrics*, 27(4), 167–172. doi:10.1002/jae.1278
- R Development Core Team. (2013). *R: A Language and Environment for Statistical Computing*. (“R Foundation for Statistical Computing,” Ed.). Vienna. doi:10.1007/978-3-540-74686-7
- Raykov, T. (1997). Scale Reliability , Cronbach ’s Coefficient Alpha , and Violations of Essential Tau- Equivalence with Fixed Congeneric Components. *Multivariate Behavioral Research*, 32(4), 329–353. doi:10.1207/s15327906mbr3204
- Raykov, T. (2001). Bias of coefficient alpha for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, 25(1), 69–76.

doi:10.1177/01466216010251005

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to Psychometric Theory*. New York: NY: Taylor & Francis.

Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, 47(5), 667–696. doi:10.1080/00273171.2012.715555

Reise, S. P., Waller, N. G., & Comrey, a L. (2000). Factor analysis and scale revision. *Psychological Assessment*, 12(3), 287–297. doi:10.1037/1040-3590.12.3.287

Revelle, W. (2015a). Alternative estimates of test reliabiity. Retrieved from <http://personality-project.org/r/html/guttman.html>

Revelle, W. (2015b). Package “psych.” Retrieved from <http://personality-project.org/r/psych-manual.pdf>

Revelle, W., & Zinbarg, R. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145–154.

Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, 26(1), 108–16. doi:10.7334/psicothema2013.260

Santisteban, C. (2009). *Principios de Psicometría*. Madrid: Síntesis.

Santisteban, C., & Alvarado, J. (2001). *Modelos Psicométricos*. Madrid: UNED.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1(2), 199–223. doi:10.1037/1082-989X.1.2.199

- Schreiber, J., Nora, A., Stage, F., Barlow, E., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–337. Retrieved from <http://www.tandfonline.com/doi/abs/10.3200/JOER.99.6.323-338>
- Shapiro, A., & ten Berge, J. M. F. (2000). The asymptotic bias of minimum trace factor analysis, with applications to the greatest lower bound to reliability. *Psychometrika*, 65(3), 413–425. doi:10.1007/BF02296154
- Sheng, Y., & Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? *Frontiers in Psychology*, 3(34), 1–13. doi:10.3389/fpsyg.2012.00034
- Sijtsma, K. (2009a). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107–120.
- Sijtsma, K. (2009b). Reliability beyond theory and into practice. *Psychometrika*, 74(1), 137–143. doi:10.1007/s11336-008-9100-1
- Sijtsma, K., & van der Ark, L. A. (2015). Conceptions of Reliability Revisited and Practical Recommendations. *Nursing Research*, 64(2), 128–136. doi:10.1097/NNR.0000000000000077
- Sočan, G. (2000). Assessment of Reliability when Test Items are not Essentially t-Equivalent. *Developments in Survey Methodology*, 15, 23–35.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72–101. doi:10.1037/h0065390
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha. *Journal*

*of Personality Assessment*, 80(1), 99–103.

Tang, W., & Cui, Y. (2009). *Internal Consistency: Do We Really Know What It Is and How to Assess it?*

Tarkkonen, L., & Vehkalahti, K. (2005). Measurement errors in multivariate measurement scales. *Journal of Multivariate Analysis*, 96(1), 172–189.  
doi:10.1016/j.jmva.2004.09.007

ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69(4), 613–625.  
doi:10.1007/BF02289858

Teo, T., & Fan, X. (2013). Coefficient Alpha and Beyond: Issues and Alternatives for Educational Research. *Asia-Pacific Education Researcher*, 22(2), 209–213.  
doi:10.1007/s40299-013-0075-z

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is Datametrics: the Test is not Reliable. *Educational and Psychological Measurement*, 60(2), 174–195.  
doi:10.1177/0013164400602002

Trianes, M. V., Blanca, M. J., Morena, L. De, & Raya, L. I. S. (2006). Un cuestionario para evaluar el clima social del centro escolar. *Psicoth*, 18(2), 272–277.

Wilcox, S., Schoffman, D. E., Dowda, M., & Sharpe, P. A. (2014). Psychometric Properties of the 8-Item English Arthritis Self-Efficacy Scale in a Diverse Sample. *Arthritis*, 2014.  
doi:10.1155/2014/385256

Wilkinson, L., & American Psychological Association Task Force on Statistical Inference,

- T. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594–604. doi:10.1037/0003-066X.54.8.594
- Woodhouse, B., & Jackson, P. H. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: II: A search procedure to locate the greatest lower bound. *Psychometrika*, 42(4), 579–591.
- Yang, Y., & Green, S. B. (2011). Coefficient Alpha: A Reliability Coefficient for the 21st Century? *Journal of Psychoeducational Assessment*, 29(4), 377–392. doi:10.1177/0734282911406668
- Zimmerman, D. W. (1975). Probability spaces, Hilbert space, and the axioms of Test Theory. *Psychometrika*, 40(3), 395–412.
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's  $\alpha$ , Revelle's  $\beta$ , and McDonald's  $\omega$ H: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133. doi:10.1007/s11336-003-0974-7
- Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. (2006). Estimating Generalizability to a Latent Variable Common to All of a Scale's Indicators: A Comparison of Estimators for  $\omega$ h. *Applied Psychological Measurement*, 30(2), 121–144. doi:10.1177/0146621605278814
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21–29. Retrieved from [http://digitalcommons.wayne.edu/jmasm/vol6/iss1/4/?utm\\_source=digitalcommons.wayne.edu%252Fjmasm%252Fvol6%252Fiss1%252F4&utm\\_medium=PDF&utm\\_campaign=](http://digitalcommons.wayne.edu/jmasm/vol6/iss1/4/?utm_source=digitalcommons.wayne.edu%252Fjmasm%252Fvol6%252Fiss1%252F4&utm_medium=PDF&utm_campaign=)



Zumbo, B. D., & Rupp, A. A. (2004). Responsible modelling of measurement data for appropriate inferences: Important advances in reliability and validity theory. In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 73–92).